Web Search as a Linguistic Tool

Adam Fourney
Microsoft Research
Redmond, Washington
adamfo@microsoft.com

Meredith Ringel Morris
Microsoft Research
Redmond, Washington
merrie@microsoft.com

Ryen W. White
Microsoft Research
Redmond, Washington
ryenw@microsoft.com

ABSTRACT

Many people rely on web search engines to check the spelling or grammatical correctness of input phrases. For example, one might search [*recurring or reoccurring*] to decide between these similar words. While language-related queries are common, they have low click-through rates, lack a strong intent signal, and are generally challenging to study. Perhaps for these reasons, they have yet to be characterized in the literature. In this paper we report the results of two surveys that investigate how, when, and why people use web search to support low-level, language-related tasks. The first survey was distributed by email, and asked participants to reflect on a recent search task. The second survey was embedded directly in search result pages, and captured information about searchers' intents in-situ. Our analysis confirms that language-related search tasks are indeed common, accounting for at least 2.7% of all queries posed by our respondents. Survey responses also reveal: (1) the range of language-related tasks people perform with search, (2) the contexts in which these tasks arise, and (3), the reasons why people elect to use web search rather than relying on traditional proofing tools (e.g., spelling and grammar checkers).

Categories and Subject Descriptors

H.3.3 [**Information storage and retrieval**]: Information search and retrieval—*Query formulation*

Keywords

language-related queries; web search; spelling; grammar

1. INTRODUCTION

People often rely on web search engines to perform a variety of low-level, language-related tasks. For example, searchers issue queries to verify the correct spelling of words, to disambiguate between homonyms, or to perform any number of similar tasks (Table 1). These linguistic uses of web search are so common and well-accepted that they have become part of our shared public consciousness. As an example, the Merriam-Webster Online Dictionary tracked the themes of the 2016 US presidential election by examining the words web searchers looked up throughout the campaign [16]. These efforts and findings were widely reported in the media [23]. Likewise, to celebrate the Scripps National Spelling Bee, Google recently revealed which words web searchers most frequently asked [*how to spell*], in each of the 50 US states [18].

Hyphenation	Deciding about hyphenation, or about joining words (e.g., follow up/follow-up).
Homonyms	Deciding between similar sounding words (e.g., affect vs effect).
Grammar	Checking if a phrase is grammatically correct (e.g., "in regard to" or "in regards to").
Spelling	Checking the spelling of a word or proper noun.
Definition	Learning the definition of an unfamiliar word or phrase.
Pronunciation	Learning how to pronounce a word or proper noun.
Thesaurus	Finding similar or opposite words (i.e., synonyms or antonyms).
Etymology	Learning the history or origin of a word or phrase.

Table 1: Eight linguistic tasks that people perform with web search engines.

In contrast, there is a surprising lack of research characterizing the phenomenon of using web search to support low-level linguistic tasks. This scarcity of literature persists despite the fact that language-related search strategies have motivated several automatic grammar-checking systems [5, 8, 15, 21, 22, 28], and frequently serve as illustrative examples in the study of positive web search abandonment[1] [1, 13, 24]. As we will show, this situation can perhaps be explained by the fact that language-related web searches have low click-through rates, lack a strong intent signal, and are generally challenging to study.

[1]Here, positive abandonment refers to the scenario where a searcher's information need is met by the contents of the search engine results page (SERP), thus eliminating the need to click on a result [4].

WWW 2017, April 3–7, 2017, Perth, Australia.
ACM 978-1-4503-4913-0/17/04.
http://dx.doi.org/10.1145/3038912.3038914

In this paper we report the findings of two surveys designed to directly study how, when, and why people rely on web search engines to support linguistic tasks. The first survey collected responses from 149 people, and asked respondents to reflect on a recent language-related search task. The second survey was embedded directly in search result pages, and captured information about the language-related queries issued by 142 individuals over a six week period. Taken together, responses to these surveys directly address the following research questions:

> **RQ1**: How often do people rely on web search engines to support linguistic tasks? How common is this behavior?
>
> **RQ2**: What specific language-related tasks do people use web search to perform? Which of these tasks can be directly addressed by the contents of SERPs?
>
> **RQ3**: In what contexts or scenarios do people use web search to support language-related tasks?
>
> **RQ4**: Why do people use generic web search engines for these tasks, rather than relying on the specialized proofing tools now embedded in most software applications and services?

In the remainder of this paper we review related work, then detail the methods and data used in our analyses. We address the four research questions in turn, then discuss the limitations and broad implications of this research.

2. BACKGROUND

The observation that people rely on web search engines to support language-related activities is not new. In 2008, while developing a classification scheme for searcher intents, Jansen et al. remarked [12]:

> Web search engines are spell checkers, thesauruses, and dictionaries. (...) People are employing search engines in new, novel, and increasingly diverse ways.

At around the same time, Jacquemont et al. [11], Yi et al. [28], and Park et al. [21] independently observed that writers often performed web searches to decide if particular phrasings were common in English documents. For example, a searcher might query [*"take a seat"*] followed by [*"take the seat"*] to determine, from document counts, which preposition is correct. This class of errors is especially common among non-native English writers [21, 28]. These anecdotal observations inspired several automated proofing systems [5, 8, 11, 15, 17, 21, 22, 28]. These systems treat the web as a normative corpus of text in a given language [19, 22], and rely on web search engines to quickly compute document frequency statistics (e.g., [5]) and to collect sample phrases (e.g., [28]). In some cases, systems also consider the query suggestions returned by the search engine (e.g., the text following "*Did you mean*", on Google) [17]. The success of these efforts demonstrates that web search engines can be leveraged to support low-level linguistic tasks. However, these papers do not directly study the day-to-day search practices that originally inspired these innovations.

In the same vein as the abovementioned work, prior research has explored using query logs to develop spelling and grammar checkers [3, 9, 14]. Again, the key property of these corpora is that correct strings generally occur more frequently than incorrect strings [3]. Moreover, modern search engines often suggest spelling corrections, and query logs maintain records of which suggestions are accepted by searchers [9]. This feedback is an important signal used to continuously improve the spell checker. Continual refinement may help explain *why* some people prefer to use web search engines over more traditional proofing tools, but these papers did not explicitly study searcher motivations or practices.

Much of what is known about language-related searches comes from research studying positive abandonment in web search [1, 13, 24]. Positive abandonment describes situations where a searcher's information needs can be directly satisfied by the contents of a search result page, therefore negating the need for users to click through to view individual search results [4]. We describe the relevant research below.

In what is perhaps the most detailed investigation of language-related search activities, Stamou and Efthimiadis asked six graduate students to label the intents and outcomes of every web search they performed over the course of one week [24]. The authors reported that one intent, "*look(ing) for linguistic information about query terms (e.g., spelling, translation, definition)*", accounted for 4.76% of all web search queries that were recorded. Moreover, 73.9% of these searches resulted in positive abandonment. While limited in scale, the research conducted by Stamou and Efthimiadis presents the first empirical evidence of how frequently language-related queries occur, and how infrequently users engage with the results of these searches. In this paper, we replicate these findings with a study that is roughly thirty times larger. Furthermore, we extend our understanding of this phenomenon by exploring language-related sub-tasks, contexts and motivations.

Additional details about language-related queries are provided by work conducted by Li et al. in [13]. An interesting dimension of their work is that it considers queries posed in three distinct locales: The United States, China and Japan. Li et al. sampled 400 abandoned web search queries from each locale, and, after manual labeling, identified dictionary and spelling-related searches as having tremendous potential for positive abandonment. Intriguingly, this potential was greatest for Japanese queries originating from mobile devices. Unfortunately, Li et al. do not provide base rates for abandonment, and do not reveal how frequently people rely on search engines to perform language-related tasks.

Chilton and Teevan also briefly discuss language-related search queries in their work investigating the impact of presenting answer cards directly in line with search results [1]. In particular, the authors studied 15 common answers categories (e.g., weather reports, phone numbers, etc.) and found that dictionary answers were among the least likely to receive user interactions. These findings reinforce the theme that language-related queries can often be addressed with no further user interaction beyond simply issuing a search query. Again, language-related queries were not the focus of that work, and the authors provided few details about the prevalence of those queries.

Finally, the topic of language-related searches also arises in Ido Guy's comparison of spoken and typed queries [10]. In this work, Guy reports that dictionary answers are triggered 1.43 times as often when people issue voice queries, as compared to when queries are entered with the keyboard (on mobile devices). In this same vein, Ong and Suplizio conducted market research on Amazon Echo users, and found that 17.6% of early adopters had issued voice queries asking the Echo for assistance in "spelling something" [20]. Together, these findings suggest that language-related queries are also common in voice search scenarios. Voice search is an interesting dimension to consider, but is ultimately out of the scope of this paper.

In summary, prior research provides several compelling clues about how people use web search to support low-level linguistic tasks. However, these details are disparate, and are buried in papers describing a range of *other* phenomena. To the best of our knowledge, there has yet to be any research directly characterizing how, when or why people use web search to support these linguistic tasks.

In this paper, we deliberately and methodically investigate the common linguistic uses of web search engines. Where possible, we replicate prior results. We also extend prior work by characterizing the specific tasks, scenarios, and motivations that underlie this phenomenon.

3. METHODS AND DATA

As we will show later in this paper, many language-related search queries consist only of a single term, and often fail to receive any clicks. These factors render language-related searches extremely difficult to study using traces or web logs alone. To this end, our work engages directly with users by means of a complementary pair of short surveys: The first survey is retrospective, and the second gathers searcher responses in-situ (Figure 1). We describe each survey below, then present the results of our analyses in the next section.

3.1 Retrospective Survey

3.1.1 *Design and Distribution*

An invitation to complete an anonymous online survey was emailed to a random sample of 2000 employees within Microsoft Corporation. The survey began by collecting demographic information. It then asked respondents to recall a specific example of a recent English-language search query that they issued to support reading, writing or another similar language-related activity. To help respondents ground their responses in a recent experience, the questionnaire included links to pages listing their recent Bing.com[2] and Google[3] searches. Provided that respondents could recall a specific example, the survey asked respondents to describe the task they were performing (Table 1), and to answer a series of follow-up questions. Each question was displayed in a multiple-choice format, and provided an option allowing respondents to write-in their own answers if they preferred.

3.1.2 *Data and Demographics*

In total, 149 respondents completed the questionnaire (response rate = 7.45%). Within this group, 110 respondents were male (73.8%), and 134 reported having a bachelor's degree or greater college education (89.9%). Respondent ages were distributed as follows: 43 participants were between the ages of 25-34, (28.9%); 53 respondents were between 35-44, (35.6%); and 39 participants were between 45-54 (26.2%).

The demographics questionnaire also asked respondents to list the language in which they were most comfortable reading and writing. Here, 145 participants reported that they preferred *reading* in English (97.3%), while 148 participants reported they preferred to *write* in English (99.3%). As such, in contrast to prior research that focused on English-language learners (e.g., [21, 28]), our work characterizes the language-related searches of those who consider themselves fluent in the English language.

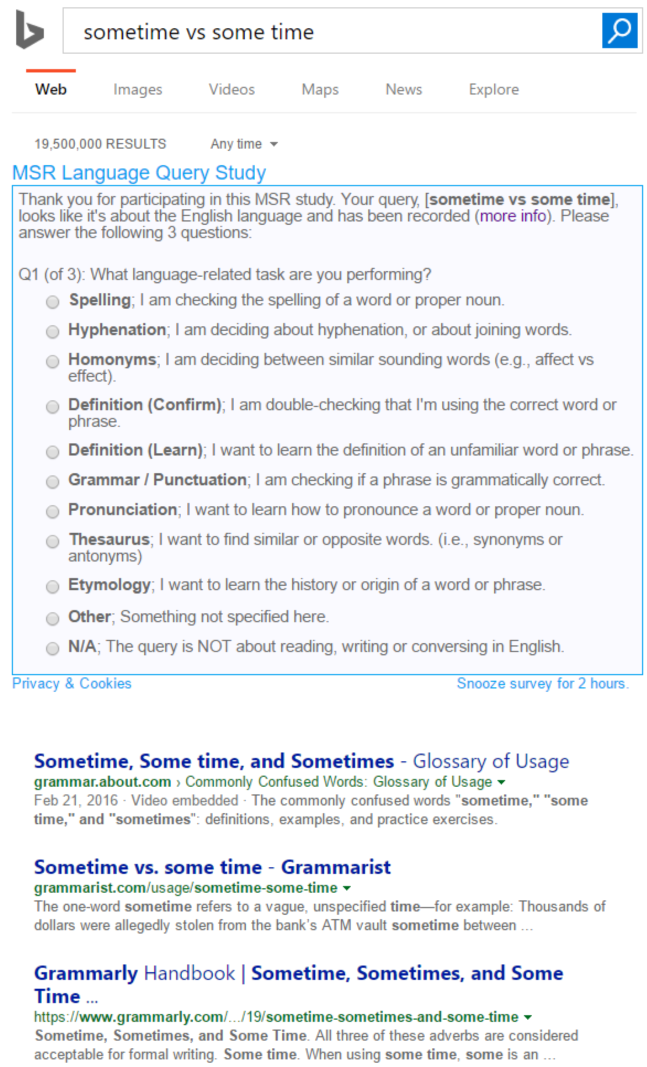

Figure 1: A screen shot of the first question posed by the in-situ survey. The in-situ survey was presented inline with search results upon detection of a potential language-related search query.

3.2 In-situ Survey

The in-situ survey (Figure 1) was implemented as a browser extension, and largely follows the design principles outlined by Diriye et al. in [4]. Given recent efforts to standardize

[2] `https://www.bing.com/profile/history`
[3] `https://history.google.com/history`

browser extension development[4], the in-situ survey extension could be installed on a range of modern browsers. We now describe: (1) the conditions under which the survey was triggered, (2) which questions were asked, and (3) what additional instrumentation data was collected.

3.2.1 Triggering the In-situ Survey

The extension monitored web searches performed on both the Bing and Google search engines. The in-situ survey was then triggered whenever a dictionary answer was present on a SERP. The survey was also displayed when any of the top three organic search results linked to one of the top 100 most frequently visited English dictionaries, thesauri, grammar guides, style guides, or other linguistic references known to the Bing.com search engine (as measured in January, 2016).

3.2.2 In-situ Survey Design

Upon detecting a potential language-related query, the browser extension injected a short survey directly in line with the search results (Figure 1). The first survey question asked participants to report the language-related task they were in the process of performing. The survey presented the same set of options as was available in the email survey, with two important distinctions: First, the survey included a *not applicable* option allowing users to flag false positive classifications. Second, the *dictionary definition* task was divided into the two sub-tasks listed below:

- Confirm definition; Double-checking that one is using the correct word or phrase.
- Learn definition; Learning the definition of an unfamiliar word or phrase.

The second in-situ survey question asked participants to explain why they elected to use web search rather than the spell-checker, grammar-checker, or dictionary that is built into desktop software (e.g., Outlook, Office, Chrome, Firefox, etc).

The third and final question asked participants if the search results page contained enough information to directly satisfy their information need.

Provided that participants answered at least one of the three questions, the survey was snoozed, and would not reappear for 30 minutes (regardless of which queries a participant issued).

3.2.3 Instrumentation

In addition to recording survey responses, the extension gathered basic information about every query the user issued, and every mouse or keyboard interaction the user performed on search results pages (similar to the instrumentation described in [4]). In instances where queries failed to satisfy the conditions for potential language-related searches, the extension logged only cryptographic hashes of the text input by searchers or displayed on the results page. This design afforded users a degree of privacy, while allowing us to compute aggregate statistics and meet our research objectives.

[4]`https://developer.mozilla.org/en-US/Add-ons/WebExtensions`

3.2.4 Deployment

As with the email survey, the in-situ survey was deployed within Microsoft Corporation. Participants were recruited via an email invitation sent to a random sample of 5000 employees occupying a variety of job roles within the company. Recipients were asked to install the browser extension for six weeks. To incentivize participation, we randomly selected one participant per week to receive a $50 gift card for an online retailer. To avoid the possibility of biasing participants with monetary rewards, we made it clear that the selection process did not depend on how frequently participants engaged with the browser extension.

3.2.5 Data and Demographics

In total, the browser extension was installed by 143 distinct individuals (response rate = 2.9%). Participants shared similar demographics to those responding to the email survey (no person participated in both studies). As noted above, participants occupied a diverse set of roles including software development, sales and marketing, legal services, supply chain engineering, and technical writing, to name a few. As before, most respondents (93%) listed English as their preferred language for both reading and for writing.

4. RESULTS

We now present the results of our two surveys. The discussion is organized so as to address our four research questions.

4.1 Frequency of Language-Related Searches

To determine how frequently people leverage web search to support low-level language-related tasks, and to answer our first research question, we analyzed the instrumentation data collected by the in-situ survey. In total, 29,211 queries were observed, and 891 (3.05%) were classified as potentially language-related (i.e., triggered a dictionary answer, or linked to a linguistic web site, as defined earlier). There was no statistically significant difference between the proportions of queries identified as language-related on Bing and Google ($z = 1.088$, $p = 0.276$). These figures represent a median of 5 weeks of search activity collected for each participant (IQR = 3). As a sanity check, we compared these rates to the rates at which the same conditions arise in the production Bing search engine, and we found a similar low single-digit value.

In response to the 891 positive classifications, the browser extension triggered the in-situ survey 690 times (the survey was suppressed 201 times as a result of the 30-minute timeout). Users responded to the survey on 381 occasions (55.2% engagement). Of those 381 responses, 48 queries were labeled as false positives. This suggests a classifier precision of 87.4%. Under the assumption that this precision applies to the 510 unlabeled positive classifications, we estimate that the extension recorded 776 ± 30 language-related searches (confidence interval of 95%). This represents between 2.56% to 2.76% of all queries observed. We note that this figure should be considered a lower bound on the proportion of language-related search activity, as our in-situ data does not allow us to estimate the false negative rate.

For one point of comparison, recall that Stamou and Efthimiadis exhaustively labeled *all queries* posed by six computer-science graduate students over a single week. They found that 4.76% of search queries were related to language [24]. Our lower-bound is compatible with this figure, and adds

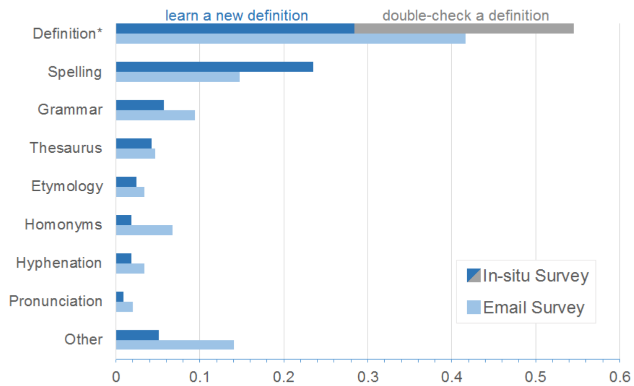

Figure 2: Bars indicate the proportions of language-related web searches ascribed to eight tasks, as labeled by respondents of the in-situ and email surveys. In the in-situ survey, the *definition* task was subdivided. For comparison purposes, these sub tasks are shown together as a stacked bar graph.

Write-in task	Responses
Double-check a definition	"*double-checking the meaning*" "*confirm the definition of a word*" "*check if a word was used correctly*" "*check that the spelling matched my definition of the word*"
Translation	"*I wanted a translation to include in a presentation*" "*Reading a site in French*" "*Kanji (translation) to go into a presentation*" "*Translation*"
Acronyms / Abbreviations	"*acronym meaning*" "*I wanted to know what the abbreviation stood for*" "*check acronym*"

Table 2: The retrospective survey allowed respondents to write in their own language-related tasks in cases where they felt the eight tasks listed in the survey failed to apply. This table lists three common tasks reported by respondents.

roughly thirty times as much evidence to support the conclusion that: language-related queries account for a small but meaningful proportion of all search traffic (about 1 in every 40 searches, in our data set).

4.2 Linguistic Tasks

Our second research question asks: What specific language-related tasks do people perform with web search? A natural follow-up question is: Which of these information needs can be directly addressed from the search results pages (i.e., without requiring users to click a result). We address both questions in turn.

4.2.1 Task Variety

Respondents from both surveys reported using web search to perform a broad range of language-related tasks. Though not directly comparable[5], we present responses together in Figure 2 to emphasize their shared trend. Notably, retrieving definitions and checking spelling are the two most common tasks, followed by checking grammar/punctuation and looking up synonyms or antonyms. In both cases, word pronunciation was the task least-frequently reported by users. Finally, we note that the homonym disambiguation task is anomalous – this task was reported nearly three times more often in the retrospective survey than was observed in the in-situ survey. We have several hypotheses that may explain this discrepancy. For example, it is possible that the in-situ survey was not sensitive to these tasks, and failed to trigger. Alternatively, it is possible that, although rare, these tasks are more memorable, and are more likely to be reported in the retrospective survey. We leave further analyses of homonym disambiguation queries to future research.

In addition to considering in-situ survey responses, we also inspected the corresponding search queries. 172 (51.7%) of language-related queries consisted of only a single query term (e.g., [*dreamt*]). Considered in isolation, these queries lack a clear intent signal. Paired with survey responses, the queries can be broken down by task. We found that spelling queries were 2.5 times as frequent in this setting. This difference is highly statistically significant (two-tailed, $z = 4.298$, $p << 0.05$). However, we were also able to identify single-term queries for each and every linguistic task we considered in the survey. This highlights why it was important that we engage with users directly rather than relying only on log data.

Finally, the retrospective survey allowed users to write in their own tasks if they so desired (appearing under the *other* category in Figure 2). Inspection of the 21 write in responses reveals a number of common themes, as depicted in Table 2. The first theme, double-checking definitions, motivated our decision to subdivide the *definition* task into two sub tasks when deploying the in-situ survey.

4.2.2 Task Support

In the previous section, we reported that respondents leveraged web search to support a broad range of linguistic tasks. In this section, we report which tasks could be accomplished directly from the SERPs. To investigate this, both surveys sought to identify instances of positive abandonment. Here, the in-situ survey simply asked respondents if they were satisfied with the search results, then monitored the SERPs for mouse clicks. Conversely, the retrospective survey asked respondents to reissue their queries and comment on the results.

71.0% of in-situ survey responses, and 72.5% of retrospective survey respondents, indicated instances of positive abandonment. These rates compare favorably with the 73.9% figure reported by Stamou and Efthimiadis in [24].

We are also able to extend prior work by leveraging the in-situ responses to break abandonment rates down across the range of linguistic tasks (Figure 3). Notably, retrieving definitions, verifying spelling, deciding about hyphenation, and distinguishing between homonyms, each resulted in positive abandonment rates above 70%. Conversely, the etymology and thesaurus tasks were positively abandoned less than 30% of the time. This low rate is statistically

[5]In the retrospective survey, respondents could only describe one search query.

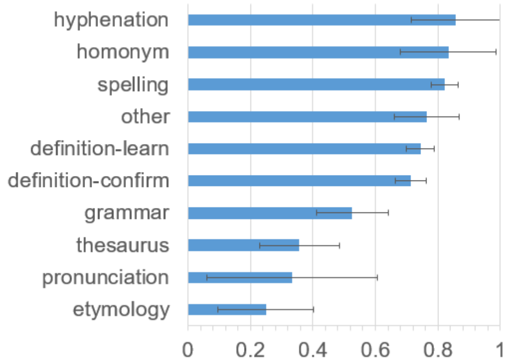

Figure 3: Bars indicate the positive abandonment rates for various language-related tasks. In these cases, respondents indicated that they were satisfied with the search results, yet did not click on any links. Error bars indicate the standard error (SE) of the sample proportions.

Writing	I wanted to include the word or phrase in something I was writing.
Proofreading	I came across the word or phrase while proofreading a document (reading with intent to edit).
Reading	I came across the word or phrase while reading (without intent to edit).
Overheard	I heard this word or phrase used in a conversation (including in a movie, on the radio, etc.).
Brainstorming	I was brainstorming (e.g., names for a product).

Table 3: Respondents to the retrospective survey were asked to indicate the scenario in which their language-related information needs arose. This table lists the options respondents could choose when answering this question. Additionally, respondents could write in their own scenario.

indistinguishable from the proportion of all 29,211 queries that experienced some form of abandonment ($p = 0.67$, and $p = 0.50$, respectively).

In summary, respondents leveraged web search to perform a wide variety of linguistic tasks. However, the queries were often under-specified and contained only a single term. Moreover, the corresponding search results were often dismissed without user interaction.

4.3 Scenarios

One advantage of the retrospective survey is that it afforded participants the opportunity to describe the scenarios in which their information needs arose (Table 3), including reporting web searches performed on mobile devices (rather than being restricted to observe only queries originating from a single desktop or laptop computer). The resultant analysis answers our third research question.

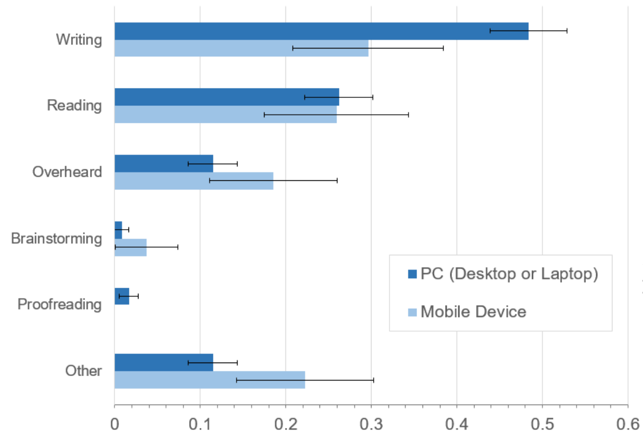

Figure 4: Respondents of the retrospective survey indicated that they performed language-related searches in a number of different scenarios, and on a variety of devices. Bars indicate the proportion of responses ascribed to each. Error bars indicate the standard error (SE) of the sample proportions.

Figure 4 summarizes the results of our analysis. Overall, 45% of respondents reported that they had issued their queries because they wanted to use a word or phrase in a document that they were in the process of writing. However, writing was much more common on computers than on mobile devices (48.4% vs. 29.6%). This difference is statistically significant (two-tailed, $z = -2.370$, $p = 0.018$), and appeals to our intuitions. Apart from writing, language-related queries arose when users were reading documents (26.2%), or after having overheard a word used in conversation (12.8%). We also found that respondents were more likely to write in their own scenarios when describing mobile searches (vs desktop searches). These responses appear as *other* in Figure 4. Some of the mobile write-in scenarios include: "*having a friendly disagreement between friends*", helping children with homework, and playing puzzle games such as Scrabble or Words With Friends.

Again, we are able to break these responses down by task type (Figure 5). As expected, the writing scenario gave rise to the most heterogeneous set of language-related information needs. Here, the most common task was checking the spelling of words, but this task accounted for only 28.4% of the responses. In contrast, the most common task performed in reading scenarios was looking up the definitions of words (comprising 69.2% of all tasks). Again, these findings appeal to our intuition, serve as a useful sanity check, and raise our confidence in the quality of the survey data.

4.4 Reasons for Using Web Search

Finally, upon detecting a language-related search query, the in-situ survey asked participants to answer the following question: Why did you just use web search for this task? (vs relying on the spell-checker, grammar-checker or dictionary built into your browser, operating system, and other desktop software). Possible responses are listed in Table 4. Figure 6 summarizes the results, and directly answers our fourth research question.

Overall, the most common explanation for relying on web search was that it was convenient – the browser was already

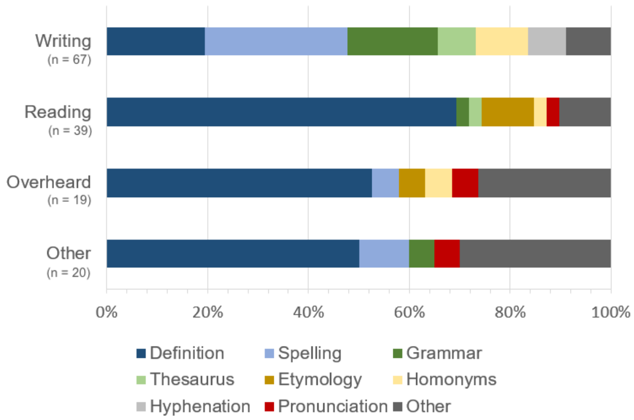

Figure 5: Bars indicate the proportions of language-related search queries ascribed to each task type across four broad scenarios. The *brainstorming* and *proofreading* scenarios are omitted due to their low number of responses.

Task support	The word or phrase is not in my spell-checker's dictionary, or the task is not supported by such tools.
Richness	Web search provides richer information (e.g., examples, images)
Trust	I am more confident that web search will provide the correct answer
Familiarity	I am more familiar with web search than with the aforementioned tools
Convenience	E.g., The browser was already open.
Other	A reason not shown here.

Table 4: Respondents of the in-situ survey were asked to indicate why they elected to use web search rather than rely on more traditional proofing tools (e.g., the spell checkers built into software applications). This table lists the set of possible responses.

open (42.6%). This was followed by task support (18.4%), richness (13.2%), trust (10.3%), and familiarity (8.7%). Figure 6 breaks down responses by task type. We find there are interesting differences across tasks. For example, *convenience* dominates the responses in cases where queries were issued to check a word's spelling (56.3% of responses), while *task support* was the most common explanation offered for the *thesaurus* task (50.0% of responses). Perhaps more surprising is that *richness* and *trust* together account for nearly a quarter (23.5%) of the linguistic uses of web search.

5. IMPLICATIONS

We now consider both the immediate and the broader implications of our findings. In the former category, we feel that our findings can serve as a useful critique on the design of the dictionary answers returned by contemporary search engines. Specifically, we note that the first two lines (and largest font sizes) of dictionary answers are dedicated to the presentation of syllabic and phonetic information (Figure 7). However, the need to learn pronunciations was very rare among our survey respondents. More common was the need to verify the correct spelling of words. We hypothesize that the syllabic and phonetic representations may interfere with this more-common task (e.g., by rendering the correct spelling more difficult to read [25], copy & paste, or recognize by sight). We recommend exploring ways to tailor these answers to a searcher's stated or implied linguistic tasks.

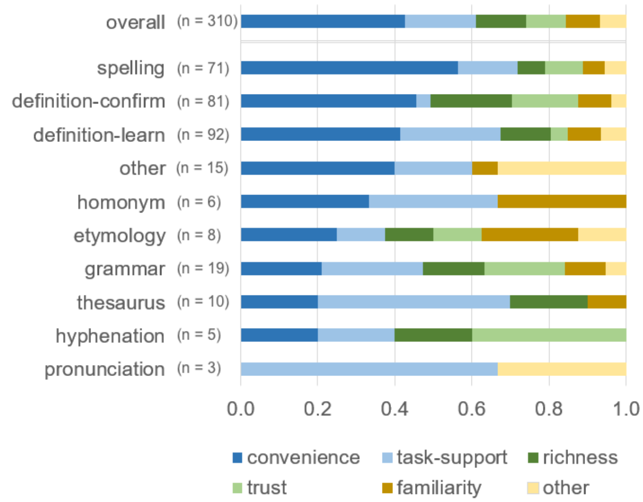

Figure 6: Colors indicate the various reasons why respondents of the in-situ survey reported relying on web search to perform language-related tasks.

Similarly, our findings have implications for the design of writing tools. *Richness* and *trust* were two of the more frequent responses people offered when asked about their use of web search in this context. Perhaps next-generation spelling and grammar checkers should consider enriching their suggestions with information retrieved from the web. As Fallman noted in [5], it can often be tremendously helpful to see and compare the number of search results returned for various spellings or phrasings of a word or sentence. Moreover, search engines and online content have evolved considerably in the fourteen years since Fallman's work was published. As such, it is worth considering other SERP features that might add richness to, and build trust in, spell checkers (e.g., photographs, Wikipedia summaries, and structured information about entities).

More generally, we feel that language-related search queries present an opportunity for search engines to work together with content creation software in service of better supporting users and their tasks [6]. For example, one could imagine a scenario where clicking on a dictionary card in a SERP results in the corresponding word being inserted into whatever document the user was currently editing. This interaction would be especially useful on mobile devices where multi-tasking is difficult, and where copy & paste gestures are cumbersome. Notably, smartphone on-screen keyboards are already beginning to include web search capabilities[6], but do not yet gracefully handle language-related queries.

Synchronizing search engines with content creation software may also benefit the underlying search services. For

[6] https://googleblog.blogspot.com/2016/05/gboard-search-gifs-emojis-keyboard.html

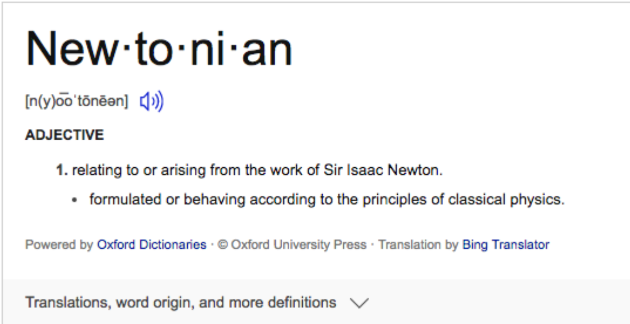

Figure 7: A screen shot of the dictionary answer returned for the query [*newtonian*] on Bing.com. The Google search engine returns a similar dictionary card for this query. This presentation is optimized to facilitate *pronunciation*, but may make it harder to *verify spelling*. Our data suggests the latter is a much more common task.

example, while searchers may not click on results when issuing language-related queries, they are likely to use the information to refine text they are composing in another application [7]. Detecting these explicit or implicit transfers of information (e.g., copy & paste events, or re-typing text found on SERPs) could serve as useful implicit relevance signals – especially in cases of positive abandonment. Likewise, if a search engine was aware that a document was being edited concurrently, it might be better able to determine which queries should trigger dictionary answers – or more generally, when language-related tasks were being performed.

To conclude, pairing search engines with content creation software, and recognizing the roles of language-related queries, affords opportunities to: help users efficiently compose text, help search engines decide when to return linguistic answers, and help evaluators measure the success of search interactions.

6. DISCUSSION AND FUTURE WORK

The survey results analyzed in this paper indicate that language-related searches are common among a group of highly-educated people who reported having good fluency in both spoken and written English, and who worked for a technology company in the United States. We caution readers against overly generalizing the results reported here. For example, it is known form prior research that both domain expertise [26] and web search expertise [27] can impact how people use search engines.

Nevertheless, it seems likely that certain demographic groups, such as people who are learning English as a non-primary language, children, and people with reading or writing disabilities (e.g., dyslexia) may benefit even more from language-related search than the group we studied. For example, English-language learners were the target audiences for many of the prior search-inspired grammar checkers mentioned earlier (e.g., [21, 28]). We also find that certain classes of language-related queries closely align to the American academic calendar (Figure 8). This raises the possibility that students or educators are among those who frequently leverage web search for linguistic purposes.

Future work studying the frequency and nature of language-related queries for these user groups may reveal differences

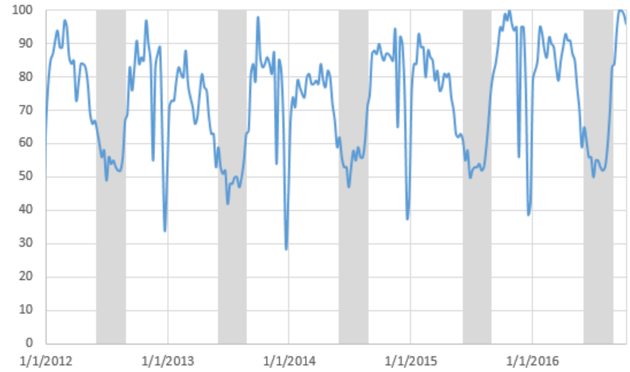

Figure 8: Relative search volume for the homonym disambiguation query [*affect effect*], from January 2012 to October 2016 (as reported by Google Trends[7]). Shaded regions indicate the months June-July. Local minima occur during the summer, over American Thanksgiving, and over Christmas break. These patterns align with the academic year, and may indicate that students or educators are among those using web search to support linguistic tasks.

from the group we studied, and may suggest specialized ranking schemes or user interfaces tailored to the needs of these audiences. For example, both Bing and Google currently return definitions from the Oxford English Dictionary, but younger students may benefit from a dictionary written at a lower reading level [2]. Likewise, English language learners may benefit from a bilingual dictionary. Finally, we note that language-related search frequency and task types may differ for searches conducted in languages besides English.

7. CONCLUSION

In this paper we reported the results of two surveys that investigate how, when, and why people use web search to support low-level, language-related tasks. Our findings confirm that language-related searches are indeed common (accounting for at least 2.7% of queries in our data set), and often result in positive abandonment. This latter property renders language-related queries difficult to study using trace logs alone. Fortunately, our survey-based methods allowed us to characterize this phenomenon in detail. We found that people perform a wide variety of linguistic tasks, but retrieving definitions and verifying spelling dominate. Likewise, we found that, for users of desktop computers, linguistic queries were most often associated with writing; however, queries were also issued while users were reading text or listening to spoken conversations. Finally, we found that convenience drove a plurality of all language-related web searches (roughly 40%), but that task-support, richness, and trust (in the results) also gave rise to many linguistic queries. Taken together, our findings paint a rich picture of this common but rarely studied use of web search engines. Future work includes expanding our results to other demographics and device contexts, as well as exploring opportunities to more tightly integrate web search engines with content creation applications.

[7] `https://www.google.com/trends/`

8. REFERENCES

[1] L. B. Chilton and J. Teevan. Addressing people's information needs directly in a web search result page. In *Proceedings of WWW '11*, pages 27–36, New York, NY, USA, 2011. ACM.

[2] K. Collins-Thompson, P. N. Bennett, R. W. White, S. de la Chica, and D. Sontag. Personalizing web search results by reading level. In *Proceedings of CIKM '11*, pages 403–412, New York, NY, USA, 2011. ACM.

[3] S. Cucerzan and E. Brill. Spelling correction as an iterative process that exploits the collective knowledge of web users. In *Proceedings of EMNLP 2004*, pages 293–300, Barcelona, Spain, July 2004. Association for Computational Linguistics.

[4] A. Diriye, R. White, G. Buscher, and S. Dumais. Leaving so soon?: Understanding and predicting web search abandonment rationales. In *Proceedings of CIKM '12*, pages 1025–1034, New York, NY, USA, 2012. ACM.

[5] D. Fallman. The penguin: Using the web as a database for descriptive and dynamic grammar and spell checking. In *Proceedings of CHI EA '02*, pages 616–617, New York, NY, USA, 2002. ACM.

[6] A. Fourney, B. Lafreniere, P. Chilana, and M. Terry. Intertwine: Creating interapplication information scent to support coordinated use of software. In *Proceedings of UIST '14*, pages 429–438, New York, NY, USA, 2014. ACM.

[7] A. Fourney and M. R. Morris. Enhancing technical Q&A forums with CiteHistory. In *Proceedings of ICWSM '13*, 2013.

[8] M. Gamon and C. Leacock. Search right and thou shalt find...: Using web queries for learner error detection. In *Proceedings of IUNLPBEA '10*, pages 37–44, Stroudsburg, PA, USA, 2010. Association for Computational Linguistics.

[9] J. Gao, X. Li, D. Micol, C. Quirk, and X. Sun. A large scale ranker-based system for search query spelling correction. In *Proceedings of COLING '10*, pages 358–366, Stroudsburg, PA, USA, 2010. Association for Computational Linguistics.

[10] I. Guy. Searching by talking: Analysis of voice queries on mobile web search. In *Proceedings of SIGIR '16*, pages 35–44, New York, NY, USA, 2016. ACM.

[11] S. Jacquemont, F. Jacquenet, and M. Sebban. Correct your text with google. In *Web Intelligence, IEEE/WIC/ACM International Conference on*, pages 170–176, Nov 2007.

[12] B. J. Jansen, D. L. Booth, and A. Spink. Determining the informational, navigational, and transactional intent of web queries. *Information Processing & Management*, 44(3):1251 – 1266, 2008.

[13] J. Li, S. Huffman, and A. Tokuda. Good abandonment in mobile and pc internet search. In *Proceedings of SIGIR '09*, pages 43–50, New York, NY, USA, 2009.

[14] M. Li, Y. Zhang, M. Zhu, and M. Zhou. Exploring distributional similarity based models for query spelling correction. In *Proceedings of ACL-44*, pages 1025–1032, Stroudsburg, PA, USA, 2006. Association for Computational Linguistics.

[15] A. D. Matthieu Hermet and S. Szpakowicz. Using the web as a linguistic resource to automatically correct lexico-syntactic errors. In *Proceedings of LREC'08*, Marrakech, Morocco, may 2008.

[16] Merriam-Webster Incorporated. Tracing the presidential election through the words Americans looked up most. `http://www.merriam-webster.com/words-at-play/trending-words-from-election-2016/`, 2016. [Online; retrieved 20-October-2016].

[17] J. Moré. A grammar checker based on web searching. *Digithum*, 0(8), 2006.

[18] J. W. Moyer. Google uncovered 'America's top spelling mistakes,' and 'diarrhea' is on the list. `https://www.washingtonpost.com/news/local/wp/2016/05/26/google-uncovered-americas-top-spelling-mistakes-and-diarrhea-is-on-the-list`, May 2016. [Online; retrieved 20-October-2016].

[19] K. A. Olsen and J. G. Williams. Spelling and grammar checking using the web as a text repository. *J. Am. Soc. Inf. Sci. Technol.*, 55(11):1020–1023, Sept. 2004.

[20] S. Ong and A. Suplizio. Unpacking the breakout success of the amazon echo. `https://www.experian.com/innovation/thought-leadership/amazon-echo-consumer-survey.jsp`, 2016. [Online; retrieved 20-October-2016].

[21] T. Park, E. Lank, P. Poupart, and M. Terry. 'Is the sky pure today?' AwkChecker: An assistive tool for detecting and correcting collocation errors. In *Proceedings of UIST '08*, pages 121–130, New York, NY, USA, 2008. ACM.

[22] J. Sjöbergh. The internet as a normative corpus: grammar checking with a search engine. Technical report, KTH Nada, Stockholm, Sweden, 2006.

[23] L. Stack. 'Braggadocious?' in Trump and Clinton debate, words that sent viewers to the dictionary. `http://www.nytimes.com/2016/09/28/us/braggadocio-in-trump-and-clinton-debate-words-that-sent-viewers-to-the-dictionary.html`, September 2016. [Online; retrieved 20-October-2016].

[24] S. Stamou and E. N. Efthimiadis. Interpreting user inactivity on search results. In *Proceedings of ECIR'2010*, pages 100–113, Berlin, Heidelberg, 2010. Springer-Verlag.

[25] M. Tinker. *Legibility of print*. Iowa State University Press, 1963.

[26] R. W. White, S. T. Dumais, and J. Teevan. Characterizing the influence of domain expertise on web search behavior. In *Proceedings of WSDM '09*, pages 132–141, New York, NY, USA, 2009. ACM.

[27] R. W. White and D. Morris. Investigating the querying and browsing behavior of advanced search engine users. In *Proceedings of SIGIR '07*, pages 255–262, New York, NY, USA, 2007. ACM.

[28] X. Yi, J. Gao, and W. B. Dolan. A web-based english proofing system for english as a second language users. In *Proceedings of IJCNLP'08*, pages 619–624, 2008.