Industry Watch

Checking in on grammar checking

R O B E R T   D A L E

Arria NLG plc
e-mail: Robert.Dale@arria.com

(*Received 9 March 2016*)

Abstract

Ten years ago, Microsoft Word's grammar checker was really the only game in town. The
software world, and the world of natural language processing, have changed a lot in that
time, so what does the grammar checker marketplace have to offer today?

---

1 An appropriate time to reflect

I am writing this piece on a very special day. It's National Grammar Day, 'observed'
(to use Wikipedia's crowdsourced choice of words) in the US on March 4th.[1]

The word 'observed' makes me think of citizens across the land going about their
business throughout the day quietly and with a certain reverence; determined, on
this day of all days, to ensure that their subjects agree with their verbs, to not their
infinitives split, and to avoid using prepositions to end their sentences with.

I can't see it, really. I suspect that, for most people, National Grammar Day
ranks some distance behind National Hug Day (January 21st) and National Cat
Day (October 29th). And, at least in Poland and Lithuania, it has to compete with
St Casimir's Day, also celebrated on March 4th. I suppose we could do a study to
see whether Polish and Lithuanian speakers have poorer grammar than Americans
on that day, but I doubt we'd find a significant difference.

So National Grammar Day might not mean all that much to most people, but
it does feel like an appropriate time to take stock of where the grammar checking
industry has got to. I last wrote a piece on commercial grammar checkers for the
Industry Watch column over 10 years ago (Dale 2004). At the time, there really was
no alternative to the grammar checker in Microsoft Word. What's changed in the
interim? And does anyone really need a grammar checker when so much content
these days consists of generated-on-a-whim tweets and SMS messages?

---

[1] See nationalgrammarday.com.

2 The evolution of grammar checking

Grammar checking software has evolved through three distinct paradigms. First-generation tools were based on simple pattern matching and string replacement, using tables of suspect strings and their corresponding corrections. For example, we might search a text for any occurrences of the string *isnt* and suggest replacing them by *isn't*. The basic technology here was pioneered by Bell Labs in the UNIX Writer's Workbench tools (Macdonald 1983) in the late 1970s and early 1980s, and was widely used in a range of more or less derivative commercial software products that appeared on the market in the early '80s. Anyone who can remember that far back might dimly recall using programs like *RightWriter* on the PC and *Grammatik* on the Mac.

Second-generation tools embodied real syntactic processing. IBM's Epistle (Heidorn *et al.* 1982) was the first really visible foray into this space, and key members of the team that built that application went on to develop the grammar checker that, to this day, resides inside Microsoft Word (Heidorn 2000). These systems rely on large rule-based descriptions of permissible syntax, in combination with a variety of techniques for detecting ungrammatical elements and posing potential corrections for those errors.

Perhaps not surprisingly, the third generation of grammar-checking software is represented by solutions that make use of statistical language models in one way or another. The most impressive of these is Google's context-aware spell checker (Whitelaw *et al.* 2009)—when you start taking context into account, the boundary between spell checking and grammar checking gets a bit fuzzy. Google's entrance into a marketplace is enough to make anyone go weak at the knees, but there are other third-party developers brave enough to explore what's possible in this space. A recent attempt that looks interesting is *Deep Grammar* (`www.deepgrammar.com`).

We might expect to find that modern grammar checkers draw on techniques from each of these three paradigms. You can get a long way using simple table lookup for common errors, so it would be daft to ignore that fact, but each generation adds the potential for further coverage and capability.

3 Today's grammar-checking marketplace

So: you're in the market for a grammar checker, and for whatever reason, you're not happy with Microsoft's offering. What's out there that you might consider?

If you want an interesting summary of the state-of-the-art with regard to current products, you might check out John Thiesmeyer's website (`www.serenity-software. com`). Thiesmeyer is unlikely to be unbiased, since he does have his own grammar-checking application to sell. But he does a pretty thorough job of comparing the capabilities of around twenty grammar checking applications, and makes some acute observations about the field generally. Of course, in an age of agile software development, with teams copying each others' good ideas as soon as they are publicly available, analyses like these will necessarily be out-of-date very quickly, but it's a good place to get an idea of the range of things these applications can and can't deal with.

One prominent aspect of Thiesmeyer's review is his disdain for what he calls 'the noisy three' grammar checkers, those being *Grammarly*, *Style Writer* and *WhiteSmoke*. He's railing against the marketing efforts of these companies. In that regard, noisy they are, and *Grammarly* is the noisiest of them all. If you search the web for grammar checkers, you're likely to find many more mentions of *Grammarly* than anything else. Those mentions seem to crop up in three kinds of places: in comparative reviews of grammar checking software,[2] as links or ads on pages that initially appear to be offering some other web-based checking tool,[3] and, perhaps most frequently, as recommendations in blog postings. *Grammarly* appears to be everywhere. And it has over six million likes on Facebook. I haven't seen any figures on market share, but I would not be surprised if it was the market leader.

This high visibility all starts to make sense when you look at *Grammarly*'s affiliate program: if you place a link to *Grammarly* on your website, you get twenty cents for each click-through registration for the free version of the product, and $20 for each registration for the premium version (which costs the user $25 a month). It's hardly surprising, then, that people are happy to talk up the capabilities of the software. Even people who don't seem to have bothered using the tool on their own copy wax lyrical about it. The following appears in a graphic on one of these sites:[4]

> About 90% Visitors Who Visit Your Site or Read Your Presentation, checks Grammer.
> First Impression is always best impression.
> Almost 85% of Traffic Is Not Converting into Leads Due to Bad Grammer.

I have no idea what impact writing like this has on sales of the product being recommended, but perhaps any publicity is good publicity.

Affiliate marketing is a standard practice across the web in all sorts of domains, so it shouldn't be a surprise that we find it here. I'm not picking on *Grammarly* here; the other members of 'the noisy three' also have affiliate programs, but it's harder to find out what their terms and conditions are, since you have to register with the sites before they are revealed. *Grammarly*'s more open approach has perhaps served them well in terms of take-up.[5]

4 Capabilities

What do you get in a modern grammar checker? In writing this piece, I informally experimented with ten of the current offerings (*ClearEdits*, *Correct English*, *Editor*, *Ginger*, *Grammarly*, *GrammarBase*, *GrammarCheck.net*, *Pro Writing Aid*, *SpellCheck-Plus*, and *Style Writer 4*) and I was universally underwhelmed. *Grammarly* and

---

[2] For example, becomeawritertoday.com/grammar-checker-review-grammarly.

[3] Such as www.grammarchecksoftware.org, www.grammarcheckeronline.info, www.grammarcheckonline.net, and www.sentencechecker.org. And probably a host of other cognate URLs I didn't check.

[4] See myreviewcoupon.com/grammarly-review. This text is as it appeared on the page on National Grammar Day, but it may have been corrected by the time you read this.

[5] Just to be clear: I'm not an affiliate of any these marketing programs. If, as a consequence of reading this piece, you decide to buy any of the mentioned products, please give five dollars to a good cause.

Ginger appear to be a little ahead of the pack, presumably because their success so far has enabled further R&D; but some of the more minor players seem stuck in an earlier decade, both in terms of the techniques used—with a heavy reliance on first-generation techniques—and in terms of delivery mode. Modern services tend to be delivered through a subscription SaaS model; some products are available as web browser plug-ins or as Word plug-ins. Some of the older packages are effectively standalone applications, and they feel very tired. The quality of the interfaces varies tremendously. The ability to do some form of grammar checking, if you screw up your eyes and ignore the actual quality, is now almost a commodity, and packages distinguish themselves by incorporating other features, such as translation, rich dictionary definitions, read-aloud functionality and personalised training technology that feeds you exercises based on the kinds of errors you make. But the grammar checking capabilities, as far as I can tell, are still somewhere behind what you get in MS Word.

5 Who needs a grammar checker?

Obviously, there are many people who might benefit from assistance with writing, ranging from professionals who write as part of their jobs (the apparent target of *StyleWriter*) to special interest groups like scriptwriters (the focus for *AutoCrit*). One group that many products appear to target are ESL students, aiming to help them improve the quality of their essays. This is also, of course, a target group for academic research in the area.[6]

A perhaps not-so-obvious market consists of men looking for dates. *Grammarly* partnered up with eHarmony to conduct a study on how spelling and grammar mistakes influence interactions on dating sites, surveying 10,000 eHarmony profiles for spelling, punctuation and grammar errors.[7] Apparently, men who properly use *whom* get thirty-one per cent more contacts from the opposite sex. Just two spelling errors on a man's profile reduces his chances of a response by fourteen per cent. On the other hand, according to this study, women make twice as many writing mistakes as men, but this doesn't impact on their chances of a positive match. So now you know.

Of course, these days a lot of communication comes in transient 140 character lumps. But even Twitter usage is not immune from grammatical scrutiny: a review of the text quality of fifty top celebrities' tweets determined that Conan O'Brien had the least number of errors, Bill Gates came a close second, and Barack Obama third.[8] Khloe Kardashian just gets into the list at tenth. If Bill was using a grammar checker, I wonder which one?

---

[6] See Dale (2014) for a review.
[7] See `http://musingsandmarvels.com/2016/02/11/from-grammarly-valentines-day-and-online-dating`.
[8] See `http://www.brit.co/celebs-grammar-day`. Yes, this was another survey carried out by *Grammarly*.

References

Dale, R. 2004. Industry watch. *Journal of Natural Language Engineering* **10**(1): 91–94.

Dale, R. 2014. Automated writing assistance. In R. Mitkov (ed.), *The Oxford Handbook of Computational Linguistics*, 2nd ed. Oxford University Press.

Heidorn, G. 2000. Intelligent writing assistance. In R. Dale, H. Moisl, and H. Somers (eds.), *A Handbook of Natural Language Processing: Techniques and Applications for the Processing of Language as Text*, New York: Marcel Dekker, pp. 181–207.

Heidorn, G., Jensen, K., Miller, L., Bird, R., and Chodorow, W. 1982. The EPISTLE text critiquing system. *IBM Systems Journal* **21**(3): 305–326.

Macdonald, N. H. 1983. The UNIX writers workbench software: rationale and design. *Bell System Technical Journal* **62**(6): 1891–1908.

Whitelaw, C., Hutchinson, B., Chung, G. Y., and Ellis, G. 2009. Using the web for language independent spellchecking and autocorrection. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, vol. 2, Stroudsburg, PA: Association for Computational Linguistics, pp. 890–899.