

Anniversary INDUSTRY WATCH

NLP commercialisation in the last 25 years

Robert Dale*

Language Technology Group

*Corresponding author. Email: rdale@language-technology.com

Abstract

The *Journal of Natural Language Engineering* is now in its 25th year. The editorial preface to the first issue emphasised that the focus of the journal was to be on the practical application of natural language processing (NLP) technologies: the time was ripe for a serious publication that helped encourage research ideas to find their way into real products. The commercialisation of NLP technologies had already started by that point, but things have advanced tremendously over the last quarter-century. So, to celebrate the journal's anniversary, we look at how commercial NLP products have developed over the last 25 years.

1. Some context

For many researchers, work in natural language processing (NLP) has a dual appeal. On the one hand, the computational modelling of language understanding or language production has often been seen as means of exploring theoretical questions in both linguistics and psycholinguistics; the general argument being that, if you can build a computational model of some phenomenon, then you have likely moved some way towards an understanding of that phenomenon. On the other hand, the scope for practical applications of NLP technologies has always been enticing: the idea that we could build truly useful computational artifacts that work with human language goes right back to the origins of the field in the early machine translation experiments of the 1950s.

However, it was in the early 1990s that commercial applications of NLP really started to flourish, pushed forward in particular by targeted research in both the USA, much of it funded by the Defense Advanced Research Projects Agency (DARPA) via programs like the Message Understanding Conferences (MUC), and Europe, via a number of large-scale forward-looking EU-funded research programs. It was against this backdrop of activity that the present journal came into being, with an explicit goal of focussing primarily on practical rather than theoretical contributions.

With the journal now in its 25th year, we have a nice excuse to look at the history of NLP commercialisation and to reflect on how far we have come in those 25 years.

2. Starting points

As well as being the year in which this journal began, 1995 was notable for a considerable number of computing-related events. It was the year in which Windows 95 and Netscape became available. Java 1.0 appeared, and JavaScript was developed. DVDs were introduced, and Sony released the PlayStation in North America. It was also the year in which NSFNet was decommissioned, removing the last restrictions on the commercialisation of the Internet. Yahoo.com, eBay.com and Amazon.com all launched in 1995. And IBM unveiled Deep Blue, the computing system that went on to beat world chess champion Garry Kasparov. Ah, those were heady days!

Table 1. Church and Rao's assessment of NLP commercialisation in 1995

	Well-understood	State-of-the-art	Forward-looking
NLP	NL interfaces	Entity extraction and generation from databases	Event extraction
Word processing	Simple string matching	Spell checkers	Grammar checkers
Information retrieval	Keyword search	NL search	Conceptual search
Machine translation	Glossary look-up	Translation memories and direct transfer	Interlingual MT

Conveniently for our present purposes, 1995 was also the year in which Ken Church and Lisa Rau published an article on the commercialisation of NLP (Church and Rau 1995). It's worth reviewing how the field appeared at that time. Church and Rau looked at developments in four areas: NLP generally, word processing, information retrieval and machine translation. In each case they attempted to characterise what constituted well-understood techniques, what counted as state-of-the-art and what was more forward-looking. Table 1 restates the content of the relevant summarising figure from that paper.

In some regards, it might seem like little has changed: although progress has been made in each of the topics identified by Church and Rao as forward looking, they are not what you would call completely solved problems, and so I suppose they remain forward looking in some regard, perhaps along the lines that true AI is whatever we haven't built yet. But in a variety of ways that we'll explore below, the field has moved forward in many dimensions, and there are NLP-based products in everyday use today that we would not have considered plausible 25 years ago.

Most notable, of course, has been the change in approach to many problems that has resulted from the widespread adoption of machine learning and statistical techniques; in the mid-1990s, there was research being done on statistical approaches, but anything that had reached product-hood was likely to be primarily rule-based in terms of its technological underpinnings. The new techniques have not only provided new solutions to old problems but have brought new problems onto the stage, resulting in product ideas we would not have thought of before.

3. The commercial landscape

For the purposes of the analysis provided here, I'll adopt a decomposition of the space of NLP applications and technologies that is a little different to that used by Church and Rau, and which corresponds to one I've used in earlier columns. That view of the world breaks down the commercial NLP space into six key areas:

Machine translation: The meaning-preserving translation of linguistic content in a source natural language into a target natural language.

Speech technologies: The conversion of a wave form corresponding to a linguistic utterance into a textual rendering of that utterance (speech recognition); or, conversely, the conversion of a textual representation of a linguistic utterance into a corresponding audio form (speech synthesis or TTS).

Dialog interfaces: Interactive interfaces that permit a user to ask a natural language question (either via speech or text) and receive an answer, possibly in the context of a longer multi-turn dialog.

Text analytics: The detection of useful content or information in textual sources, typically either via classification or via extraction.

Natural language generation: The generation of linguistic output from some underlying non-linguistic representation of information.

Writing assistance: Technologies that embody some knowledge of language for the purposes of improving the quality of human-authored linguistic content.

There are other ways one could decompose the field, of course. I find this way of looking at things to be useful because it corresponds reasonably well to the set of self-assigned categories that vendors in the space tend to use. It's arguable that speech technologies no longer represent a separate category, having been somewhat absorbed into the broader area of dialog interfaces. However, at least back in the 1990s, there was a clear separation between the technologies used for turning speech into text on the one hand, and for working out what to do with the resulting text on the other hand.

We'll look at each of these six categories in turn in the remainder of this article.

4. Machine translation

As noted earlier, the first experiments in machine translation began in the 1950s, achieving widespread visibility with the now-famous Georgetown-IBM Russian-to-English translation demonstration in 1954 (Hutchins 2004). Commercial providers came onto the scene in the late 1960s and early 1970s, and in fact one of the earliest companies in the space, Systran (www.systransoft.com) – founded over 50 years ago in 1968 – still exists as a leading provider of translation tools and services.

Early MT systems were purely symbolic, based on large collections of hand-written rules that would convert forms in one language into another language. PC versions of commercial MT software began to appear in the early 1990s. It would be later in the 1990s before web-based translation appeared (recall that the web only became publicly available in 1991), with AltaVista's Babel Fish translator in 1997 – built using Systran's technology. So, 25 years ago, commercial MT was a rule-based endeavour, and much of the commercial focus was on workbenches embodying functionalities, such as translation memories, that increased productivity. Although IBM published research on statistical machine translation (SMT) in the late 1980s, it was 2007 before web-based translation became statistical, with the introduction of Google Translate.

Today, the translation industry is in a state of flux. Sometimes-rough SMT is considered good enough for some purposes, and there's no doubt that the use of deep learning techniques has improved results. But machine translation still falls some way short of human quality and accuracy, so there's an increasing number of hybrid solutions that attempt to combine the scalability of machine translation with the precision and accuracy of human translators.^a The widespread availability and low cost of machine translation have led to a great deal of soul searching in the wider translation industry: each time the quality of machine translation improves, human translators have to reconsider what it is that they bring to the party.^b This is an existential anxiety that we can expect to see play out for a few more years to come.

But back in 1995, who would have thought that you would be able to point your phone's camera at a Japanese menu and see an English translation on the screen?^c That was the stuff of science fiction.

5. Speech technologies

The speech community embraced statistical techniques long before the rest of the NLP world; in particular, the adoption of HMMs in the 1970s was a major breakthrough that led to considerable improvements in recognition performance. But progress in speech recognition was still held up by the limitations in memory and computing power of the time. In the late 1970s, IBM had technology that could recognise a vocabulary of 1000 words – provided each word was spoken as an isolated unit with a pause before the next. By the 1990s, the technology had improved

^aSee, for example, <https://lilt.com>.

^bSee, for example, <https://slator.com/sponsored-content/how-augmented-translation-will-redefine-the-value-of-translators/>.

^cSee <https://www.cnet.com/news/gaijins-rejoice-ar-app-translates-japanese-menus-to-english/>.

and computational power had increased to the extent that speech recognition on the PC desktop was feasible. Dragon Dictate, released in 1990 for DOS-based computers, cost \$9000 but could still only recognise isolated words. Dragon Naturally Speaking, which could recognise up to 100 words per minute of continuous speech (once you had trained it to the specifics of your voice for 45 minutes), appeared in July 1997 and was matched later that same year by IBM's release of a competing product called ViaVoice^d.

The mid-1990s, then, were early days for the commercial deployment of speech recognition. More than any other area, it's here that the astounding advances that have been made over the last 25 years are most obvious: contrast the isolated word recognition of the early 1990s with the accuracy of recognition we have now come to expect from Alexa, Siri and Google Assistant.

Commercially available speech synthesis technology has also undergone massive improvements over the last 25 years. Many computer operating systems have included speech synthesisers since the early 1990s. The robotic DECTalk^e – made famous as the voice of Stephen Hawking – was introduced in the mid-1980s, but speech synthesis technology had not changed much by 1995. The adoption of concatenative synthesis in the 1990s led to much more natural-sounding synthesised output, but still with audible glitches and a sense that the output wasn't quite convincing. The last few years have seen significant developments, though, thanks to the application of deep learning techniques. The high quality of today's deployed speech synthesis is most evident if you listen to Google Duplex's dialog contributions, complete with hesitations and fillers,^f or Alexa's recently introduced newscaster voice.^g

6. Dialog interfaces

I use the term 'dialog interfaces' to cover a number of different technologies:

- Early 1960s' and 1970s' experiments in text-based conversational interaction, exemplified by Joseph Weizenbaum's Eliza (Weizenbaum 1966).
- Text-based natural language database interfaces of the kind that were a focus of attention in the 1970s and 1980s, allowing users to type questions like 'How many sales people are earning more than me?' rather than having to formulate an SQL query (Hendrix 1986).
- Voice dialog systems of the 1990s, where finite-state dialog modelling was combined with grammar-based speech recognition in telephony-based applications.^h
- Today's virtual assistants, such as Siri, Alexa and Google Assistant.
- Today's text-based chatbots, found on many websites and messaging apps, displaying a wide range of capabilities (or, often, lack of capability).

These might seem all like very different kinds of applications, but they all have in common that they attempt to support some form of natural language dialog. In the technology as it has surfaced for most everyday users today, that dialog is often limited to one-shot question and answer sessions ('Hey Alexa, what's the time?'). However, a number of vendors in the last couple of years have made much of the need to become more conversational, maintaining a dialog over a number of turns, and consequently having to deal with dialog phenomena such as conversational context and anaphora. Consequently, many of the lessons that were learned by the interactive voice

^dIf you want an authoritative history of the development of speech technology, you can't go past Roberto Pieraccini's *The Voice in the Machine* (Pieraccini 2012).

^eCheck out the examples at www.youtube.com/watch?v=8pewe2gPDk4.

^fSee <https://ai.googleblog.com/2018/05/duplex-ai-system-for-natural-conversation.html>.

^gSee <https://techcrunch.com/2019/01/16/alexa-gets-a-professional-newscaster-voice-for-reading-the-days-news>.

^hSee www.voicexml.org.

recognition developers of the 1990s and 2000s are being relearned by today's chatbot developers. Bruce Balentine's important contributions on this topic are as relevant today (and should be on the bookshelf of every dialog or chatbot developer), as they were 10–20 years ago (Balentine 1999; 2007).

Mark Twain, accused of almost every quotable quote, is reputed to have said that 'History doesn't repeat itself but it often rhymes'. So, while newbie dialog designers are repeating the mistakes of their forebears from 20 years ago, the difference today is that speech recognition actually works sufficiently well that there is relatively little distinction between text-based interfaces and speech-based interfaces. In the 1980s and 1990s, it was not unusual for NLP work on dialogue to simply assume the existence of perfect speech recognition, a problem to be solved by some other team in the building; but the high recognition error rates of the time meant that these systems would be fundamentally unviable if let loose on spoken language.

But we have come a long way since then. In 1992, AT&T launched its voice recognition service for call routing, demonstrating that a useful application could be developed even with an extremely limited vocabulary of five terms.ⁱ The DARPA Spoken Language Systems program of the early 1990s, and particularly the Air Travel Information System (ATIS) shared task, demonstrated the impressive results that could be achieved in the labs and encouraged the beginnings of the interactive voice recognition industry. The soon-to-be heavyweights of the telephony-based IVR industry, Nuance and SpeechWorks, were both founded in the mid-1990s off the back of the progress being made in the DARPA ATIS task; by the late 1990s, an increasing number of commercially useful applications were being developed, combining, as noted above, finite-state dialog modelling and grammar-based speech recognition to manage and limit speech recognition errors.

Today, we still haven't solved the problems thrown up by prolonged naturalistic dialog. But, even given its narrow domains of application, it's hard not to be impressed by the dialogic capabilities demonstrated by Google Duplex.^j

7. Text analytics

I use the term 'text analytics' to cover a wide range of classification and extraction processes typically applied to documents. The DARPA-funded MUC conferences were key to defining and developing this area of technology, with an early initial focus on named entity recognition. By the mid-1990s, commercial entities began to appear that aimed to commercialise these approaches. Only a small number of these still exist as separate companies; the only two that I am aware of are Basis (www.basistech.com), founded in 1995, and NetOwl (www.netowl.com), founded in 1996. A number of other early players, however, have been acquired and are still active under other guise: in particular, ClearForest, founded in 1998, now forms part of the Thomson Reuters text analytics solution, branded as Open Calais (www.opencalais.com).

If you compare these older providers with many of the newer entrants to the space, the maturity of their products, particularly in terms of breadth of functionality, is apparent. But the older players have not rested on their laurels: both Basis and Open Calais appear to be keeping up to date, introducing deep learning techniques just as the newer competitors take these technologies as a starting point.

Perhaps the most notable development in the last 25 years has been the emergence of sentiment analysis as a key technology; so much so that some text analytics companies, such as Clarabridge (www.clarabridge.com), have been reborn as providers of customer experience analytics.

As with everything else, deep learning has made in-roads into the text analytics space, and a considerable number of new entrants proudly talk of their use of these new techniques as a way

ⁱThe application was used to collect and bill-to-third-party long-distance calls, and recognised five distinct responses to a recorded voice prompt: 'calling card', 'collect', 'third-party', 'person-to-person' and 'operator'.

^jSee <https://www.youtube.com/watch?v=znNe4pMCsD4>.

to distinguish themselves from the old guard. It's less clear whether these approaches are actually delivering better results than those of the more established players, who do of course have the benefit of having tuned their techniques over the intervening decades. It's interesting to note that, back in 1995, Church and Rao saw event extraction as the forward-looking topic for this area: it remains a challenge for the few commercial text analytics providers who choose to take it on. And in another instance of history rhyming, gazetteers – that mainstay of the earliest approaches to named entity recognition – are back, but in the richer form of knowledge graphs built from sources like DBpedia. Turns out string matching was a good idea in 1995, and it's still a good idea today.

8. Natural language generation

In 1995, there was virtually no commercial NLG activity. The only company I'm aware of as being active in the space, CoGenTex (<http://cogentex.com>), had been founded in 1990, and survived mostly on government contracts. But things are very different today; the last 5–10 years have seen a growing number of companies advertising themselves as being providers of NLG. Narrative Science (<https://narrativescience.com/>) and Automated Insights (<https://automatedinsights.com>) are the most visible players here, but there's also Arria NLG (www.arria.com), Ax Semantics (www.ax-semantics.com) and Yseop (<https://yseop.com/>) amongst many others. As far as I can tell, most of these companies appear to be using some variant of 'smart templates' to generate routine texts such as financial reports or data-driven news stories. A number of these vendors are also producing plug-ins for business intelligence tools like Tableau and Qlik, adding narrative content to the high-quality visualisations these tools already produce.^k There's little evidence that commercial NLG offerings are leveraging the richer linguistic concepts, like aggregation and referring expression generation, that have been developed in the research community; my sense is that the use cases being explored so far don't warrant the use of these ideas. This of course may change as increasingly sophisticated applications are attempted.

Commercial applications of more unconventional forms of NLG – at least from the perspective of a traditional NLG research agenda – have also appeared. Just to give two examples: Phrasee (<https://phrasee.co>) generates optimised push messages and subject lines for emails; and, productising recent advances in image captioning, Microsoft's Seeing AI (www.microsoft.com/en-us/seeing-ai) helps those with impaired vision by producing descriptions for images. The latter in particular wasn't even on the horizon in 1995.

9. Writing assistance

In 1995, automated writing assistance consisted of spelling checkers, style checkers and very simple grammar checkers based on pattern matching. Word Perfect, a major competitor to Microsoft Word in the early 1990s, licensed and subsequently acquired a product called Grammatik; MS Word licensed a competing tool called CorreCText. Like all the available grammar checkers of the period, these tools were inspired by the simple string matching approaches used in the Unix Writers Workbench. We had to wait until 1997 for Microsoft to release its own syntax-based grammar checker as part of Microsoft Word, using technology developed by Karen Jensen and George Heidorn, who had previously been responsible for IBM's Epistle grammar checking system: papers were published on Epistle in the early 1980s (Heidorn *et al.* 1982), but it was syntax-based grammar checking in Word that put NLP technology on millions of desktops 14 years later.

^kSee, for example, <https://www.tableau.com/about/blog/2016/8/bringing-power-natural-language-tableau-58325>.

By the mid-1990s, Word Perfect was in decline (although it is still available today), and Microsoft appeared to have won the war for the desktop. Word's monopoly meant that there was little incentive to further develop the grammar-checking capability, and a very high barrier to entry for new players. Consequently, not much happened commercially on this front until the Grammarly grammar and style checker appeared in 2009. In 2017, Grammarly claimed to have 6.9 million daily active users,^l most of whom use the service for free. This is a tiny user base compared to Microsoft's 1.2 billion Office users in 2016,^m but of course not everyone uses Word's grammar checker; truly comparable figures are not available. As it happens, in 2017 Microsoft started rolling out improvements to its grammar checker;ⁿ so, after years of stasis, we might hope that the arrival of Grammarly as a serious competitor will breathe new life into advances in this area. My own view is that we'll see some significant product enhancements in coming years that combine ideas from NLG with the provision of writing assistance.

10. Summing up

So there you go: a whirlwind tour of how commercial NLP has advanced during the lifetime of this journal. It's clear that the NLP industry has come a long way in that time. Back in 1995, relatively few people outside the field would even know how to expand the acronym 'NLP', and any random individual might be just as likely to assume it referred to neurolinguistic programming. Today, almost anyone with a smart phone has had direct experience of using NLP technologies via Siri or Google's Assistant, and the term is widely understood.

In one way, our expectations have grown immeasurably. We are increasingly comfortable pitching random questions to our smart speakers. We treat machine translation, via its web-based incarnations, as an on-tap commodity. But at the same time – and I find this a fascinating development – our expectations are also in a way more realistic. We are not particularly upset if Alexa misunderstands our question or Google Assistant gives a seemingly spurious answer; we accept that it's par for the course that these technologies get things wrong on occasion. Similarly, we know that we should treat the results of machine translation with some caution, and weird translations no longer warrant the humorous attention they used to generate. We appear to have reached a point where, despite being imperfect, the technology is sufficiently good to be useful in everyday life. Through a combination of ubiquitous exposure and being just good enough, NLP has become an accepted part of our day-to-day experience. Viewed from where we were 25 years ago, that's truly impressive, and this journal has played its part in that journey.

If you'd like an easy way to keep up with the key developments in the commercial NLP world, you can do that by signing up for *This Week in NLP*, a short and snappy weekly newsletter published each Friday, at www.language-technology.com/blog.

References

- Kenneth W. Church and Lisa F. Rau.** (1995). Commercial applications of natural language processing. *Communications of the ACM* 38, 71–79.
- Hutchins W.J.** (2004). The Georgetown-IBM experiment demonstrated in January 1954. In Frederking R.E. and Taylor K.B. (eds), *Machine Translation: From Real Users to Research: AMTA 2004*, Lecture Notes in Computer Science, Vol. 3265, Berlin: Springer.
- Roberto Pieraccini.** (2012). *The Voice in the Machine*, Cambridge, MA: MIT Press.

^l<https://producthabits.com/how-grammarly-quietly-grew-its-way-to-7-million-daily-users>

^m<https://www.windowcentral.com/there-are-now-12-billion-office-users-60-million-office-365-commercial-customers>

ⁿ<https://redmondmag.com/articles/2017/06/21/making-word-grammar-checker-better.aspx>

- Joseph Weizenbaum.** (1966). ELIZA – a computer program for the study of natural language communication between man and machine. *Communications of the ACM* **9**, 36–45.
- Gary G. Hendrix.** (1986). Bringing natural language processing to the microcomputer market: the story of Q&A. In *Proceedings of the 24th Annual Meeting of the Association for Computational Linguistics, ACL*.
- Bruce Balentine.** (1999). *How to Build a Speech Recognition Application*, Denton, TX: Enterprise Integration Group.
- Bruce Balentine.** (2007). *It's Better to Be a Good Machine Than a Bad Person*, Colorado Springs, CO: ICMI Press.
- Heidorn G.E., Jensen K., Miller L.A., Byrd R.J. and Chodorow M.S.** (1982). The EPISTLE text-critiquing system. *IBM Systems Journal* **21**, 305–326.