

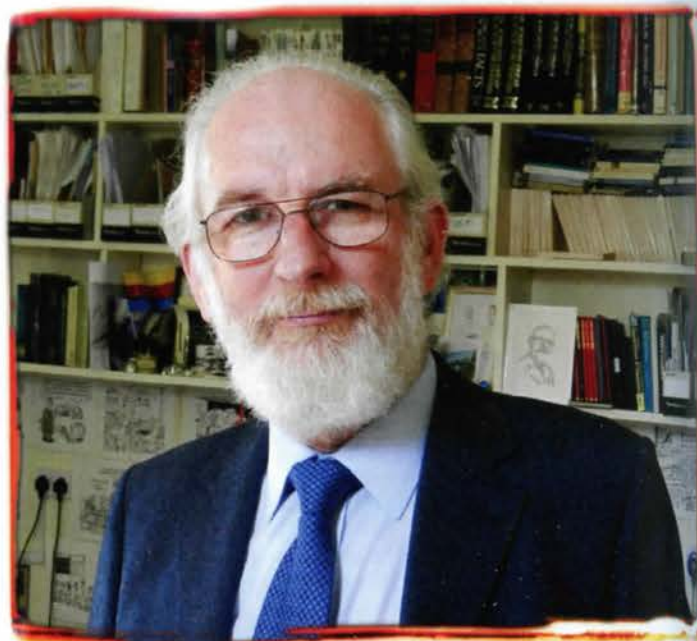
THE INTERVIEW:

Professor David Crystal

On bridging the vocabulary gap, the failure of traditional dictionaries, and semantic targeting

David Crystal is honorary professor of linguistics at the University of Bangor, and works from his home in Holyhead, North Wales, as a writer, editor, lecturer, and broadcaster. He read English at University College London, specialized in English language studies, then joined academic life as a lecturer in linguistics, first at Bangor, then at Reading, where he became professor of linguistics. He lives online at www.davidcrystal.com. Apart from the new edition of CEEL, the main project this year has been the completion and maintenance of the new edition of the Shakespeare's Words website www.shakespeareswords.com. In addition he is also, involved in the preparation of John Bradburne's enormous poetic output www.johnbradburnepoems.com for presentation to the Vatican (Bradburne's cause for beatification is progressing this year).

Helena Lustenberger, ETAS Publications Chair, interviewed Professor Crystal for *The ETAS Journal*. The conversation was wide ranging and included current trends in teaching vocabulary with specific regard to linguistics.



Helena Lustenberger: What is your recent or current focus on the linguistic study of vocabulary? Has the emphasis changed over the last 30 years?

David Crystal: There's a curious irony about the study of vocabulary. It's unquestionably the largest task a language learner has to face – I've called it the Everest of language learning – and yet, until recently, it received very little study. The point applies equally to L1 and L2 learners.

Vocabulary was hardly mentioned in the various UK government reports on language in the school curriculum – grammar got all the attention – and whether the children were learning English or a foreign language, they faced the same situation. Teachers would say, "If the kids are exposed to enough language, they'll just pick it up. And anyway, they've got a dictionary to help them."

But it seems many kids don't just "pick it up". There was a lot of talk in the UK press earlier this year about the "vocabulary gap" which is holding students back and leading to lower exam grades.

In L2 texts, the vocabulary is treated more explicitly, often in the form of lists after a chapter, but the organisation of the chapters is usually grammar related, and the vocabulary is seen only as a necessary appendage – a glossed set of words to be learned by heart. There is little system in the approach.

The first big change in recent years has been to bring vocabulary more into centre stage and to recognize that it needs the same kind of systematic study – selection, grading, grouping, etc. – that is normal in grammar and pronunciation. This emphasis was reinforced by the arrival of semantics as a major branch of linguistics during the 1970s – a seminal text was John Lyons' two-volume *Semantics* (CUP 1977) – and a programme of research that followed in which vocabulary came to be analysed using fresh perspectives.

Two points are especially important. [First,] words don't exist in isolation, but in pairs, groups, families, and we learn new words by seeing how they relate to each other. The approach introduced the notion of a semantic field (e.g. furniture, colour, animals, kinship)

and identified the various kinds of sense relation that link words (e.g. synonyms, antonyms, hyponyms).

In child language acquisition it was observed that children learn words by noting the relations between them.

"Don't touch that tap," says the parent. "That's the hot tap. You can touch the cold one, not the hot one."

Hot—cold, learned together. The sense relation is antonymy – opposites.

"What's a daffodil?" asks the child. "It's a kind of flower," says the parent. "And is that a flower too?" asks the child. "Yes, darling, that's a tulip."

Flower—daffodil/tulip, learned together. The sense relation here is hyponymy, the relationship of inclusion. An X is a kind of Y.

Which leads to the second point: that dictionaries are useless when it comes to developing an awareness of the sense relations between words. "Aunt" is under letter A; "uncle" is under letter U. Hundreds of pages apart. Words that belong together (and that should be learned together) are separated by the arbitrary nature of the alphabet.

Some modern dictionaries now try to get around this – such as the *Longman Lexicon*, which groups words into semantic fields and has an alphabetical index at the back to aid look-up.

HL: The *Longman Lexicon* does indeed provide a wealth of vocabulary grouped in semantic sets, for example "colours" and the entry for "red" gives frequency of use in spoken and written English (S1, W1), but these are more suited to fairly proficient speakers of English who are less likely to become overwhelmed by the sheer amount of information. What important information does a dictionary entry aimed at L2 A1-B1 learners need to contain?

DC: I've seen the *Lexicon* used with beginners. It's a reference work, after all, not a coursebook, so teachers can dip in and select

whatever they want. And I never cease to be amazed at the interest shown by beginners in particular areas of the lexicon – wanting to know as many words as possible in relation to a hobby, for instance.

Semantic structure exists regardless of the level of language acquisition. In mother-tongue acquisition, we see little ones of 18 months to two years acquiring basic semantic sets, and parents instinctively present them with sense relations as a means of developing their vocabulary. A “big car” is opposed to a “little car”. A “red car” is distinguished from a “blue car”.

An overextension of a word (e.g. using “elephant” to mean “animal”, so the child might call a cow an elephant) is corrected using relevant features. That’s a genuine example. The child pointed to a picture of a cow and called it an elephant. The parent replied, “That’s not an elephant. An elephant has got a big nose.” And then added a synonym, “It’s called a trunk.” And then added more semantic features, “A cow goes moo,” and so on. This is teaching sense relations.

So what are these sense relations? The examples I’ve given show:

Synonyms, where two words are very similar in meaning (never totally identical, of course) – trunk and nose

Antonyms

Hyponyms – a cow is an animal

Incompatible terms – red, blue, green, etc., are colours

Parts and wholes – an elephant has a trunk

There are other sense relations, such as series (e.g. the months of the year) and hierarchies (e.g. military ranks). Dictionaries sometimes include these, but usually as an encyclopaedic appendix.

The same point applies to incompatible sets, such as the instruments of the orchestra. All woodwind instruments form a set, as do the strings, the brass, and so on. People usually think of such groupings as part of encyclopaedic knowledge – nothing to do with language. But there is a linguistic point to be made: I can say that, “clarinets, oboes, and bassoons are all woodwind instruments”, even if I don’t know how to define the difference between a clarinet and an oboe.

Remember too that each of these general headings hides a number of variations. Opposites, for instance, are of several different kinds. There are opposites like “large” and “small”, or “happy” and “sad” which are capable of comparison (very happy, smaller). These are often called gradable antonyms.

Then there are opposites like “single” and “married”, or “alive” and “dead”, which are ungradable (we don’t say very single, not so married). These are often called complementary terms.

And then there are opposites where the two items are mutually dependent: you can’t have one without the other: “buy” and “sell”, or “wife” and “husband”. These are sometimes called converse terms.

But whatever we call them, learners need to recognize that these words work in different ways. And they also need to appreciate that there are many opposites where there isn’t a single obvious pairing. “awkward, clumsy, ungainly...” and the like are clearly the opposite of “skilful, adroit, deft...” and the like, but it isn’t possible to say that any one of these is the exact antonym of the other.

HL: One of the problems learners of English have is spelling. After hundreds of years of attempts to standardise spelling in dictionaries, the advent of modern forms of communication such as texting is leading to more “creative” spelling (e.g. “msg”, “thx”). Increased exposure to varieties of English through the Internet leads to awareness of different ways of spelling, such as “skilful” as “skillful” in American English, possibly resulting in uncertainty about correct usage. What is your view of such developments?

DC: I wouldn’t be at all concerned about text-messaging spellings. There was a great deal of over-reaction when textisms were first noticed in the early 2000s. People claimed that youngsters were filling their texts with abbreviations. In fact only about 10 per cent of the words in most texts were ever abbreviated, and most of the ones that were frequent (such as C for “see” and U for “you”) had

been in the language for a long time. IOU, for instance, is first recorded in 1618!

But textisms seem to have had their day. The advent of smartphones with predictive text has significantly reduced the number of nonstandard spellings, and amongst young people textisms are no longer considered as “cool” as they used to be. When I go into schools and talk to A-level groups (16-17-year-olds) I often ask to see a collection of their texts, as it gives me something to analyse with them. Ten years ago, there would be plenty of textisms. Now there are hardly any. One lad told me, “I stopped abbreviating when my dad started.” When adults take over young people’s language usages, they are definitely no longer cool.

But your second point is an important one. Yes, the Internet is exposing us to spelling variation as never before, especially the differences between American and British norms.

The trend has been one-way. If you plot spelling change over time (easily done these days with Google Ngrams), you see a predictable increase of American spellings in British settings, even in high-level publishing. My Cambridge University Press Encyclopedias are spelled thus, not Encyclopaedias.

And different publishing houses these days opt for different norms in spelling, hyphenation, and capitalization.

Even within British English there are alternatives, as a glance at any dictionary will show. Is it flowerpot, flower-pot, or flower pot? All three will be found and considered acceptable. The important thing is to choose one of the variants and be consistent.

In a school setting, it’s very worthwhile for the staff to agree on a spelling policy – a house style – to avoid confusing students when they are correcting their writing. At the same time, they need to point out that, whichever spelling they choose to use, they will encounter variation online, in the press, and elsewhere. Something like 12 per cent of the words in an English dictionary have spelling variation (-ise or -ize, ae or e, Moon or moon, etc.), and this increases to over 20 per cent when proper names are included: how on earth does one spell – to choose just one variant – “Tchaikovsky”? Students have a similar problem in their own language, of course.

“I wouldn't be at all concerned about text-messaging spellings... textisms seem to have had their day. The advent of smartphones with predictive text has significantly reduced the number of nonstandard spellings, and amongst young people textisms are no longer considered as “cool” as they used to be.”

The most interesting thing, to my mind, is the force the Internet is exerting to simplify spelling. English spelling is notoriously irregular, thanks to a thousand years of different influences on the system. All attempts at spelling reform have failed, apart from Webster’s for American English. Proposals for simplifying spelling are made quite often, but as no two reformers can ever agree on what is the optimal system, none have been implemented.

On the Internet, however, there are signs of regularization. A few years ago, the only spelling of “rhubarb” would have been with the “h” – a late medieval addition to the spelling of this word, based on Classical etymology. Given the pronunciation, that “h” shouldn’t be there, but if we were to leave it out in traditional written English, a copy-editor or proofreader would correct it to the standard form.

However, for most of the Internet there are no copy-editors silently correcting spellings, and so we find “rubarb” steadily increasing its frequency. A decade ago there were just a few thousand rubarb hits online. Today, a Google search tells me that there are 301 000. So, in 50 years time...? I’m sure rubarb and rhubarb

will become acceptable alternatives, just as judgment and judgement are today. And then maybe only the h-less form will survive.

HL: You have been involved in so many projects in the field of English language research over the years. Which area fascinates you most and which area would you like to turn your attention to in the future?

DC: Well, that's a consequence of the way the field has developed, over the years. Think back to the 1980s. Who could have predicted that a major new field of language study would develop in the 1990s, as a result of the Internet? Previously, we had a clear contrast between spoken and written English. Then along came electronic English, and a question immediately arose: is the linguistic character of the new medium going to be like speech, or like writing, or something in between?

The terminology was blurred: we have "written" email "conversations", for example. I wasn't intending to become an Internet researcher, but one day someone asked me if there was an introductory book on the language of computer-mediated communication. There wasn't, and so I thought to write one – enough research having been published by then to enable me to stand on a lot of shoulders – and the result was *Language and the Internet*.

Then one thing led to another, and in the late '90s I found myself applying this background to the world of Internet advertising. It started – as these things often do – with a phone call, when an advertising agency executive phoned me to ask if I could help with a problem. It seems that a CNN report of a street stabbing in Chicago was accompanied by an ad, which said, "buy your knives here".

It was, of course, an ad for cutlery, but the primitive software made no distinction between "knife equals weapon" and "knife equals kitchen tool". How could such embarrassment be avoided, I was asked.

The answer lay, once again, in semantics. Clearly, if software does not take polysemy into account – the fact that virtually all the items in a dictionary have more than one sense – it will constantly generate problems like the knife one. So – to cut the story short – the outcome was a 15-year research project, called "semantic targeting", to develop software that would guarantee the appropriate (and sensitive) placement of ads alongside web pages. The products were called iSense and Sitescreen. Fascinating – and a totally unexpected development within applied linguistics.

So you see my problem in answering your question? What will the next phone call be? The last one was in 2004, when a director at Shakespeare's Globe called to ask if I could help in reconstructing the pronunciation of Shakespeare's time for a production of *Romeo and Juliet*.

That led to another long research project – and this is still on-going, as the impact and appeal of that first production has led to many other plays being performed in OP (original pronunciation), especially in the USA, as the old accent is in many ways closer to American English than RP is.

As I write, 15 of Shakespeare's plays have been performed in OP, and each one has brought to light fresh insights – new puns are heard, rhymes work that are inexact in modern English, and a totally new phonaesthetic gives a fresh appreciation of many lines.

So my main interest at the moment is exploring OP in relation to other plays – and not just those by Shakespeare – and texts of the period. And in other periods too.

Another research team is reconstructing St Paul's Cathedral, as it would have been in the early 17th century, and wanting to hear how the liturgy and sermons would have sounded in the 1620s. That has led to an OP project later than Shakespeare – the voices of John Donne and Lancelot Andrewes. And the musical world is also just as interested in finding out about OP when singing the madrigals of the time. I think I'll be staying in this world for quite a while. But you never know...

HL: Latin gave us the basis for the classification of words used in the English language into parts of speech, a system widely adopted by grammarians and lexicographers from the days of early Modern English and widely used from the 18th century. How does a contemporary linguist view this system of classification? Are there other systems that have been considered? What is the current state of prescriptive versus descriptive lexicography?

“So – to cut the story short – the outcome was a 15-year research project, called “semantic targeting”, to develop software that would guarantee the appropriate (and sensitive) placement of ads alongside web pages.”

DC: One of the most noticeable developments in English lexicography over the past few decades has been the increased amount of grammatical information within dictionaries, especially those written for an L2 readership. It makes good sense. Words need to be observed in context for their meaning to become clear, which means putting them into sentences, or constructions within sentences. Word classification is part of this grammatical perspective, and it has been a routine feature of dictionaries, as you say, from the earliest publications – for example, Dr Johnson has an outline of English grammar in the preliminary sections of his 1755 Dictionary.

The Latinate system works to some extent, but it can't explain the whole of English grammar – there were no articles in Latin, for instance. I talk about the differences between Latin and English grammar in my *Making Sense: the Glamorous Story of English Grammar*, so I won't repeat that here, other than to say that any approach that tries to illuminate word usage needs to adopt a much more detailed approach than the basic Latin classification allows. Users need to know about transitive, intransitive, and ditransitive verbs for instance, or countable and uncountable nouns. These are important features of English, as anyone knows who has puzzled over such contrasts as, "I like cake" and "I want a cake".

So yes, other systems are now in use, based on whatever descriptive grammar the lexicographers have chosen. The process remains resolutely descriptive. In the Longman series, for example, the grammatical perspective derives from the works produced by Randolph Quirk and his colleagues (such as *A Comprehensive Grammar of the English Language*). In a way, there's nothing new about this in ELT: a chapter in a textbook on, say "countability", will simultaneously teach the grammatical point at the same time as illustrating it from relevant vocabulary. But my impression is that we are still some way from a systematic integration of grammar and lexicon in teaching materials.

References:

- English This Way 1: Nadir Kitap, Macmillan, 1964
Semantics: John Lyons, CUP 1977
Longman Lexicon of Contemporary English, Tom McArthur, Longman, 1981
The Cambridge Encyclopedia of the English Language, David Crystal, CUP, 2003
Language and the Internet: David Crystal, CUP, 2006
The Oxford Dictionary of Original Shakespearean Pronunciation, David Crystal, OUP, 2016
Guardian: 'Behemoth, bully, thief: how the English language is taking over the planet', Jacob Mikanowski, 27/7/2018
Tages-Anzeiger: 'World shut your mouth', Jean-Martin Büttner, 31/7/2018
A Dictionary of the English Language, Samuel Johnson, 1755
Making Sense: The Glamorous Study of the English Language: David Crystal, OUP 2007
A Comprehensive Grammar of the English Language: Randolph Quirk (et al.), Longman, 1985



Helena Lustenberger is the Publications Chair on the Executive Committee of ETAS. Apart from this interview, she was also the coordinating editor in this issue of Focus: Vocabulary, along with Hannah McCulloch.