

Then and Now: Competence and Performance in 35 Years of Lexicography

B. T. Sue Atkins

ITRI, University of Brighton and Lexicography MasterClass
sue.atkins@itri.brighton.ac.uk and <http://www.lexmasterclass.com>

Abstract

This paper describes the process of writing a bilingual English-French dictionary in the late 1960s, and contrasts it with a similar task in the early 21st century, with all the benefits of a large text corpus and sophisticated query tools. The lexicography focuses on the entries for *cook* and *cooking*, designed for use both by an encoding English speaker and a decoding French speaker. Lexicographic evidence for the new entries comes from a 100-million-word corpus of British English, a small corpus of French for the words *cuire*, *cuisine*, *cuisiner*, *cuisinier*, and *cuisson*, and their inflected forms, and a small parallel corpus of English and French texts. The use of KWIC concordancing, the Word Sketch program and the FrameNet database is described in detail, together with the problems of equivalence encountered. An account is given of the way in which these problems are tackled and the dictionary entry is drafted.

Dictionaries exist... to provide a series of hints and associations connecting the unknown with the known.

So runs Bolinger's dictum, quoted by Patrick Hanks in his inspiring address to the last Euralex Congress [Hanks 2000]. However, about thirty-five years ago I began writing a dictionary whose *raison d'être* was – although I couldn't have said so then – to provide a series of hints *connecting the known with the unknown*. Eleven years and fifty-odd colleagues later, this had become the Collins-Robert English-French Dictionary (CREFD), now in its fifth edition and still going strong. It has latterly benefited from the linguistic resources of HarperCollins' Bank of English, and from an editorial eye more knowledgeable than my own, yet still I see in its entries the lingering ghosts of those I wrote in my first stumbling years as a lexicographer.

In this paper I shall consider how practical lexicography has changed over the past thirty-five or so years, and set these changes in the context of a handful of closely-related English-French entries. I shall not concern myself with the great changes which the computer has brought to the consultation of the dictionaries, but solely with the writing of them. I shall look first at how the entries were written in 1967, and then at how they might be written in 2002. Any such comparison must include (as well as consideration of changes in the language itself):

- (i) the technical skill and linguistic knowledge of the lexicographer;
- (ii) the instructions or guidelines given for the production of the dictionary;
- (iii) the sources of evidence on which lexicographic decisions are based;
- (iv) any reference works available for consultation;
- (v) aids to production of text; and

(vi) the type of entry being compiled.

In this case, the first of these variables has changed radically, and the last remains the same: an English into French entry for an 'active' dictionary, i.e. one written for the encoding anglophone. Variables (ii) to (v) concern us here. I propose to focus on two specific entries, *cook* and *cooking*, to discover whether using present-day resources results in changes in the entries.

Then

In 1967, in collaboration with my French colleague Marie-Noëlle Lamy, I wrote the entries shown in Figure 1 for the first edition of the CREFD, published in 1978. For this period, the parameters of comparison listed in (i) to (vi) above are simple to report. Neither she nor I had any experience of writing dictionaries, but we had both suffered at the hands of the bilingual dictionaries we had used as language students, and were determined to do better. We learned by trial and error, and it is to Collins's eternal credit that we were given scope to do so. When I began work in 1966, the French-English half of the dictionary had already been compiled (this early text was later scrapped), and the publishers were recruiting part-time home-workers to complete the other half. I was the eighth candidate interviewed that morning, and went home with a list of words beginning with H, and the brief to write bilingual entries for them all. No other instructions were given but I was encouraged to look at other dictionaries¹ which I had to supply myself, and make my own decisions. I started to collect citations in a small card-index box. Mistrusting the other bilingual dictionaries I enrolled a local francophone linguist as informant, and six months later returned the completed 'H' entries, only to discover that A-G (and of course I-Z) still awaited compiling, the first seven lexicographers having variously fallen by the wayside. I chose C as my next task, purely on the grounds that it contained a high proportion of words of Latin origin (with initial *con-*, *contra-* etc.) which I reckoned would make for speedy compiling. The principle of a bilingual editor pair working on the text was also established at that time, and eventually we were proud to claim, on publication, that all the English and all the French in the dictionary text had been written by native speakers. This principle was maintained by Collins on most of their subsequent major bilingual dictionaries. A year later, when C was finished, alarmed by the snail's pace at which the project was progressing, Collins agreed that I should write "Compilers' Instructions" so that others could join the team, and the first version of the Style Guide came into being; over the years, enriched by many editors, it grew to 280 pages or so, and covered the lexicography of both directional halves of the dictionary.

Method of working

In writing the early English-French entries, our *modus operandi* was briefly as follows: I worked from a draft headword list supplied by the publishers. I read my reference dictionaries and tried to think up all the different constructions in which my headword might be found in each of its senses. I pencilled out a draft entry, over-rich in English examples, with suggestions for translations here and there, and notes explaining the reasoning behind the material and the way it was presented. Lamy supplied the missing French translations, often adapting the entry to include an equivalent I had not allowed for, or suggesting another example to clarify a point. We discussed the entry by telephone (living as we did in different parts of the country), she guarding the interests of the decoding francophone user and I those of the encoding anglophone reader. One of us copied out the finished entry legibly,

underlining in red for primary bold type, green for secondary bold and black for italics. Our only aids to text production were pen and paper. The dictionary text was mailed weekly in manuscript sections to the publishers, where it was typed, and indeed set in stone, since before the days of photocopiers it was never possible to check back to see what we had written in an earlier section, far less to edit it, unless some serious error came to light. We did get a chance of authors' corrections at galley proof stage but almost every correction – or so memory tells me – had to be argued for individually.

The entries

The 1967 entries are shown in Figure 1. There are a number of points of lexicographic technique which none of us would want to repeat – the treatment of compound nouns buried out of alphabetical order within a larger entry, the ubiquitous *fig* label, the ambiguous splitting of examples with *or*, and so on – but the then-and-now comparison will focus only on the content.

<p>cook [...] 1 <i>n</i> cuisinier <i>m</i>, ière <i>f</i>. she is a good ~ elle est bonne cuisinière, elle fait bien la cuisine; to be head or chief ~ and bottle-washer* (in a household) servir de bonne à tout faire; (elsewhere) être le factotum. 2 <i>cpd</i>: cookbook livre <i>m</i> de cuisine; (<i>Mil</i>, <i>Naut</i>) cookhouse cuisine <i>f</i>; (<i>US</i>) cookout repas <i>m</i> (cuit) en plein air. 3 <i>vt</i> (a) <i>food</i> (faire) cuire. (<i>fig</i>) to ~ sb's goose* faire son affaire à qn, régler son compte à qn; (b) (<i>Brit*</i>: <i>falsify</i>) <i>accounts, books</i> truquer, maquiller.</p>	<p>4 <i>vi</i> [<i>food</i>] cuire; [<i>person</i>] faire la cuisine, cuisiner. she ~s well elle fait bien la cuisine, elle cuisine bien; what's ~ing?* qu'est-ce qui se mijote?*</p> <p>◆cook up* <i>vt sep story, excuse</i> inventer, fabriquer.</p> <p>cooking [...] 1 <i>n</i> cuisine <i>f</i> (activité). plain / French ~ cuisine bourgeoise / française. 2 <i>cpd utensils</i> de cuisine; <i>apples, chocolate</i> à cuire. cooking foil papier <i>m</i> d'aluminium; cooking salt gros sel, sel de cuisine.</p>
---	---

Figure 1: The entries in the CREFD first edition

Now

In 2002, much has changed. Lexicographic training has created a pool of skilled and adaptable editors, linguistically aware and able to compile entries speedily and efficiently; dictionary projects follow detailed Style Guides, which enshrine editorial decisions made at the design stage of the dictionary, and are nowadays often consulted on line. With the advent of the computer in the 80s, many publishers installed sophisticated dictionary writing systems. These lead to a more rational approach to compiling (say, in lexical sets); they reduce the potential for errors and inconsistency by automating some routine tasks, removing a good deal of the drudgery along the way; they allow the lexicographers to look at already compiled text and to benefit from previous work by the team; they give the managing editors the opportunity of tailoring work packages to the skills and needs of the team members; they facilitate the editing process, reducing the steps in the text flow from initial editor to printed entry; and they monitor the timing and text length according to the project schedule.

And, more importantly, the computer opened the way to the lexicographers' corpus. The availability – from the early 1980s onwards – of large text corpora completely transformed

our work. In this comparison I shall focus on the various sources of lexicographic evidence now available, and review the 1967 entries in the light of what is found there.

Aids to bilingual lexicography in 2002

The source-language editor drafting an English-French bilingual dictionary entry will normally begin with an analysis (see [Atkins 1993] for a discussion of the analysis and synthesis stages of lexicography); this analysis should describe in as much detail as possible the behaviour of the English headword-lexeme² without any regard to a target language.³ At this stage, therefore, the only corpus required is an English corpus; later in the process, when the target language appears, other corpora will be helpful: a corpus of current French is essential, and parallel English-French corpora could be useful.

Drafting the framework

From the very rich analysis a 'framework' for the bilingual entry is extracted; much of the information in the framework is for internal use, and will not appear in the eventual entry. Potential dictionary senses are mapped out and standard information (e.g. parts of speech) inserted. Related multi-word expressions (idioms, compounds, phrasal verbs etc.) are included, together with other elements of the microstructure (e.g. labelling of register, style, domain etc.). The source-language editor will offer possible example sentences, and may make suggestions about translations. Notes may be added about the scope of specific source-language vocabulary items, possible faux amis or other translation problems, in an attempt to guide the target-language editor towards the best equivalents. This framework, rich in examples of usage, then goes to the target-language editor. Her responsibility is to propose equivalents, suggest amendments, and make sure that the entry does not mislead the French speaker, who has of course entirely different linguistic preconceptions from those of the English-speaking user.

Using KWIC Concordances

The first task of today's lexicographer, therefore, is to analyse the data: to discover and record significant facts about the headword, from the evidence of its behaviour available in a general corpus of current English. (The data we shall be using for *cook* is from the 100-million-word British National Corpus⁴.) The usual practice is to scan a few hundred randomly-selected concordance lines, and try to see a pattern emerging of the senses of the word. Within these senses we then look further for structured information, including (a) the constructions in which the headword participates; (b) the words with which it co-occurs most significantly (its 'collocates'); (c) any multiword expressions in which it is found; and (d) other aspects of linguistic behaviour such as register, stylistic, regional or pragmatic variation.

Our first view of the headword in the corpus is usually in the form of Key Word In Context (KWIC) lines, as shown in Figure 2. In this example, no distinction is made between noun and verb forms, although when we know that the word belongs to more than one word class it is possible to select concordances for each one separately, by using the part-of-speech tags in the corpus; this can have the effect, however, of blurring correspondences between verbs and their nominalizations.

Figure 2 shows a simple KWIC screen, produced by the Wordsmith Tools Concord program⁵ with the lines sorted on the words to the right of the keyword: there are of course other programs which allow more sophisticated searches, such as XKWIC, which is part of the Corpus Workbench developed by the IMS at Universität Stuttgart, Germany, but these tend to be used in research projects rather than commercial publishing, where time is of the essence. From the few lines shown in Figure 2, we can already glean a lot of information about *cook*, but we have to spend a good deal of time thinking about it all, even though the 14 lines in our sample represent only 0.2% of the corpus examples. We note that the word has both noun and verb uses, and in both of these word classes it seems to be polysemous. As a noun, it can refer to a profession (*Mrs James, the cook, looked comfortable*) and is also used when talking about someone's cooking skills (*I'm not much of a cook*). As a verb, it can be used transitively (*Bella cooked lunch*) and intransitively (*the ox heart cooking in the oven*). There may be a case for two senses in the transitive verb (is *cooking ham* the same sense as *cooking breakfast?*), and it occurs with a benefactive indirect object (*cooked me a man-sized breakfast*) which can also be expressed with the preposition *for* (*cooked a nice meal for Bina*).

I like	Cooking	
these are meats and poultry (raw and	Cooked)
None of them except Mrs James, the	Cook	, looked comfortable, Katherine mused.
She was totally independent, able to	Cook	, wash, look after herself and her home
she would 've gone into the kitchen to	cook	A nice meal for Bina.
a	cook	Able to produce basic dishes
I'm not much of a	cook	But it 's fun, I like doing it
he 's drawing two salaries and has me	cooking	For him.
piles of food which students attempt to	cook	For themselves to save money,
He peered into the sack and produced a	cooked	Ham
Alice	cooked	Herself eggs, drank tea, and ...
she suddenly remembered the ox heart	cooking	In the oven for Ethel
Toma made fresh coffee and	cooked	me a man-sized breakfast of bacon ...
She was sweeping the yard while Bella	cooked	The lunch and sang to the baby.

Figure 2: Some KWIC concordances for *cook*

Discovering a word's grammatical, semantic and combinatorial properties requires painstaking analysis of the data. However, there are 6750 instances of the lexeme *cook* in the BNC, too many for a lexicographer to read carefully in the time available, and mostly we make do with random sampling. Moreover, reading through hundreds of corpus lines means that important points are bound to be missed. Indeed, some types of information can never be revealed by scanning KWIC lines. For instance, the points I picked up in the concordances contain nothing about collocates, yet these are of great interest, particularly if the dictionary is one for learners of the language, as is of course the case in bilingual lexicography for the encoding user.

Statistical programs are necessary to produce information on collocate patternings: Wordsmith has some collocate functionality (showing clusters of words – e.g. *know how to, to cook for, as well as* – in the context of *cook*; offering lists of wordforms with frequency statistics according to their position vis-à-vis the keyword; indicating the distribution of the

keyword in the various texts in the corpus), but I want to look at the question of collocates through the lens of a different program devised specifically with lexicographers in mind, namely Adam Kilgarriff and David Tugwell’s Word Sketches⁶, shown in Figure 3.

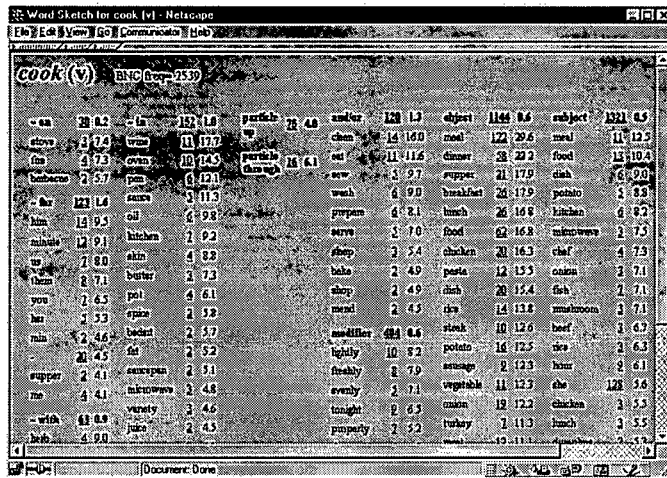


Figure 3: Part of the Word Sketch for the verb *cook*

Using Word Sketches

The Word Sketch program [Kilgarriff & Tugwell 2001] combines information of two types: grammatical relations in the corpus, and statistically significant frequencies of cooccurrence. The screen shows ordered lists of significant grammatical relations, listing each in order of salience, with the count of corpus instances; clicking on the number of instances retrieves a selection of the actual corpus examples illustrating this pattern. This means that for the verb *cook*, seen in Figure 3, the construction with the particle *up* is highly salient (rating 4.8, where the norm is 1.0) and clicking on 75 produces corpus sentences of which the first few, together with the unique corpus address, are shown in Figure 4.

5845576	If the universe consisted of just the elements	Cooked	up in the big bang , then...
6275223	how the chemical elements were	Cooked	up in the turmoil of creation
8157260	Thinking about eating, and	Cooking	up plans for new diets
28722307	Instead , it	Cooked	up a... deal with Du Pont that was to...

Figure 4: Some corpus examples of *cook* + particle *up*

Using the Word Sketch enables us to get a fix on the verb *cook* much faster than with the simple KWIC concordances. The senses of processing something (*cooking ham*) and that of preparing something (*cooking breakfast*) show up clearly in the list of objects of the verb (in Figure 3, under the heading ‘object **1144 0.6**’); the list of subjects to the right of it reminds me that food cooks (*meal, food, dish, potato*) and that people (*chef, she, I, you*) cook food; and the presence of nouns such as *meal, food* and *chicken* in both lists alerts me to the fact that this verb participates in the causative-inchoative alternation (see [Atkins et al. 1988] for a discussion of this in relation to the verb *bake*). Also I notice – what I might have

overlooked in a simple reading of KWIC lines – that appliances (*microwave*) and utensils can also be said to cook things. The list of prepositional objects following the high-salience **for** (*him, us, them, you* etc.) leaves me in no doubt of the existence of a benefactive, while – unlike the simple KWIC lines – the Word Sketch draws my attention to the possible importance to the act of cooking of the objects of the preposition *in*, and shows me at least three different lexical sets of nouns filling that slot: (1) *wine, sauce, oil* etc. (2) *oven, pan, pot* etc., and (3) *kitchen, bedsit* etc. How relevant are these to the bilingual entry? I make a note of this, to reflect on later. There is the intriguing list under the “**and/or**” heading, whose salience figure (1.3) shows me that *cook* is a verb which is found more often than the mean coordinated with another verb (*clean, eat, sew, wash* etc.). It is impossible to dig such information out of simple concordance lines, but it is something that I’ll remember when drafting the bilingual entry (how do you translate *cook* in *She can cook and clean and sew?*). Finally, the presence of *goose* at the foot of the list of objects of the verb (not visible in Figure 3) reminds me of the phrase *to cook someone’s goose*, a familiar but infrequent idiom of which there were no examples in my randomly selected set of 1,000 or so KWIC concordances.

The Word Sketch summary, combined with the ease of seeing the actual corpus examples it is based on, transforms the corpus-querying part of the analysis process: it radically reduces the time it takes to get an overview of the behaviour of the lexeme, to plan the senses that are likely to figure in the analysis entry, to look more closely at certain types of collocation in an effort to note everything of value, and to select sentences to exemplify points and possibly later to serve as examples in the dictionary entry. But even with such a tool, the lexicographer still has to disentangle the senses of polysemous words (many of them much more complex than *cook*), and cannot be sure of covering all the essential facts about the word. For assistance in these areas, we have to look to FrameNet.

Using FrameNet

The FrameNet project⁷, now in its second three-year phase, is currently building an online lexical resource, based on frame semantics and supported by corpus evidence, documenting the range of syntactic and semantic combinatory possibilities (the valence) of a word in each of its senses, through the manual annotation of example sentences. The database created during Phase 1 can be queried on the web.⁸ The project is headed by Charles J. Fillmore, and follows the credo expressed in [Fillmore 1995]:

The proper way to describe a word is to identify the grammatical constructions in which it participates and to characterize all of the obligatory and optional types of companions (complements, modifiers, adjuncts, etc.) which the word can have in such constructions, in so far as the occurrence of such accompanying elements is dependent in some way on the meaning of the word being described.

This information is exactly what lexicographers need during the analysis stage of the process.

FrameNet is of course one resource among many: its authority rests on its sound theoretical foundation and the accuracy with which the database reflects the facts of language as evidenced in the corpus. The work in FrameNet (unlike standard dictionary compiling) is

theory-informed, as well as data-driven. In the FrameNet entry there is, for each word sense, a table showing the various ways in which semantic roles, or 'frame elements', are syntactically expressed in the context of the target word. For each pattern there are supporting example sentences from the corpus, and the relevant definition from the Concise Oxford Dictionary is displayed (this last in order to locate the sense of the word for the human reader, for the entry is designed for use both by people and in a computer lexicon). Like a thesaurus, it groups words according to the semantic frames in which they participate, and work has begun on detailing the relations of semantic frames to each other. To quote the FrameNet website:

A frame is an intuitive construct that allows us to formalize the links between semantics and syntax in the results of lexical analysis. Semantic frames are schematic representations of situations involving various participants, props, and other conceptual roles, each of which is a frame element. The semantic arguments of a predicating word correspond to the frame elements of the frame, or frames associated with that word.

The objective of the FrameNet lexicographers is to record for each lexical unit (LU), or dictionary sense, every possible significant construction in which the target is found in the corpus, together with one or more corpus sentences in which the construction occurs. They aim at recording all the constructions necessary to a grammatical expression of the complex semantic range of the target word.

The FrameNet database contains three LUs for *cook* (so far – the entry is not yet complete), which may be roughly described as follows:

LU-1. heat and change the state of some foodstuff

(cook the onions, the onions were cooking)

LU-2. create a dish or a meal by doing that (also *cook up*)

(she cooked him breakfast, she likes cooking)

LU-3. invent (also *cook up*)

(...spend hours cooking up exercise programmes)

It will eventually also contain at least a fourth:

LU-4. alter (financial statements) with devious intent

(accused him of cooking the sales returns)

The first two senses (LU-1 and LU-2) will concern us in this paper.

The Apply_heat Frame

The Apply_heat frame is defined in the FrameNet manual as follows (the names of frame elements are in small capitals):

The COOK applies heat to FOOD. Heat may be applied at a certain TEMPERATURE and for a certain DURATION. A COOKING INSTRUMENT (generally indicated by a locative phrase) may also be specified. Some cooking methods involve the use of a MEDIUM by which heat is transferred to the food. In some sentences more than one phrase may contain a food frame element. When one of these phrases is less grammatically prominent than the other, *i.e.* in a prepositional phrase, it is marked FOOD2 and the other is marked FOOD1.

The verbs *fry*, *bake* and *stew* all belong to this frame. Figure 5 shows how some of these frame elements may be instantiated in the context of these verbs:

<i>Sally</i> COOK	<i>fried</i>	<i>an egg</i> FOOD	<i>in butter.</i> MEDIUM
<i>Joe</i> COOK	<i>baked</i>	<i>the cookies</i> FOOD	<i>in the oven.</i> COOKINSTR
<i>Ellen</i> COOK	<i>stewed</i>	<i>the lamb shanks</i> FOOD1	<i>with tomatoes and garlic.</i> Food2

Figure 5: Elements in the Apply_heat frame

FrameNet Annotation

Within each lexical entry (the description of one lexical unit) every example sentence selected from the corpus is fully annotated with three types of information:

- (1) the name of the frame element, or semantic role within the frame;
- (2) the type of phrase to which the annotation tag is assigned; and
- (3) the grammatical function of that phrase in relation to the target word.

Thus the sentence *Cook the meat in a saucepan over a high heat until browned* would be tagged as in Figure 6, where frame element names are shown above and grammatical information below their lexical realizations. This example is from the first LU in the lexical entry; it belongs to the Apply_heat frame (LU-1 above), as opposed to the Cook_creation frame (LU-2 above)⁹.

TARGET	FOOD	CONTAINER	TEMPERATURE	DURATION
<i>Cook</i>	<i>The meat</i>	<i>in a saucepan</i>	<i>over a high heat</i>	<i>Until browned.</i>
	NP Obj	PP Comp	PP Comp	Ssub Comp

Figure 6: Sentence showing FrameNet annotation

The annotation tags record information which is then stored in the database. For the example in Figure 6, this is:

- *cook* is the target word (the fact that it is a verb is recorded elsewhere)
- *the meat*, realizing the frame element FOOD, is a noun phrase functioning as the object of the target verb;
- *in a saucepan*, the frame element CONTAINER, is a prepositional phrase functioning as a complement of the target verb

and so on for the other two frame elements, TEMPERATURE and DURATION.

The abstract information contained in this annotation is summarized in Figure 7; in FrameNet terms this constitutes a valence pattern.

TARGET	FOOD	CONTAINER	TEMPERATURE	DURATION
	NP Obj	PP Comp	PP Comp	Ssub Comp

Figure 7: A valence pattern for *cook* in the Apply_heat frame

The complete description of this LU includes the full set of 18 valence patterns found in the corpus; these constitute the verb's valence. Figure 8 shows part of the list.

1	5 <u>exx</u>	COOK	FOOD	MEDIUM	
2	2 <u>exx</u>	COOK	FOOD	DURATION	
		--	NP Obj	PPComp	
3	1 <u>exx</u>	COOK	FOOD	MEDIUM	DURATION
		--	NP Obj	PP Comp	PPComp

Figure 8: Part of the valence of *cook* in the Apply_heat frame

Each valence pattern is exemplified in one or more corpus sentences which are called up by clicking on the 'exx' number, which indicates how many examples of the pattern have been annotated. The three valence patterns shown in Figure 8 are illustrated by the corpus sentences below:

1. She cooked the hash browns in oil.
2. Cook the duck for a further hour.
3. Cook pasta in boiling salted water for 10-12 minutes.

(In the last two sentences, the frame element COOK – the person cooking – is noted as being implicitly present, as subject of the imperative verb, but unrealized.)

Checking our first analyses and intuitions against the FrameNet entries for *cook* gives us some guidance on sense division. FrameNet offers us a structured set of facts about the lexeme-headword: these facts are already marshalled into LUs, although of course – depending on the requirements of the dictionary – we are at liberty to rework this sense division. FrameNet is however a great advance over most lexicographic resources, in that we can start from some idea of possible senses, and be fairly sure to find in the LU entry a summary of the essential elements of meaning, and the essential constructions, from which we may select what we want to include in our own dictionary entry. The FrameNet entry also offers us a direct route back into the corpus, for examples of usage: combined with the great number of corpus examples available via the Word Sketches, we will rarely need to scan KWIC lines at all. We have all we need to enable us to compile a thorough, even an exhaustive, account of the behaviour of our headword in our corpus. If the framework is correctly compiled, it should serve as a launching pad for any type of dictionary entry. A good framework should be dictionary-neutral. (This is in an ideal world, of course: in the rough and tumble of commercial publishing we never have time to create a full set of frameworks before launching on the dictionary proper.) The framework entry which results from this analysis is now ready to function as a launch-pad for our draft bilingual entry, as well as a source of information for the target-language editor, whose responsibility it is to supply the target-language items which go to make up half of the entry.

Drafting the source language entry for a bilingual dictionary

Our task as source-language editor is now to select from the wealth of facts at our disposal those most appropriate to the dictionary being compiled. We are writing a bilingual English-into-French dictionary entry, one which will not only help (passive) French speakers to understand an English text, but tell (active) English speakers enough to let them use the foreign language correctly. We must first work with the target-language editor, to tease out all the different little problems of equivalence which our headword presents in various contexts – until we do that we cannot know what information the English speaker will need. So we set about creating a very full draft: normally this is at least three times as long as the eventual entry. Already, on the basis of the framework, we must make ‘keep or lose’ decisions about grammatical constructions, collocates, example phrases and so on. At this stage we try to keep those items which are essential, those which are most likely to be useful if there is enough space, and those that have to be included in order not to mislead either the English or the French speaker. We then bundle it all up into a rough draft of an entry, already divided into senses – which may change before the entry becomes final – and complete with notes and suggestions for the target-language editor, who has to provide the equivalences throughout the entry.

Adding the target language material

Unlike English, there is no French national corpus available at low cost to researchers and lexicographers; nor is there (to the best of my knowledge) any structured corpus information like the Word Sketches or the FrameNet database to help us select the target language material for the *cook* and *cooking* entries. With the help of many colleagues, and by trawling the web, I assembled over 3,000 sentences containing *cuire*, *cuisine*, *cuisinier*, and *cuisson*, and their various inflected forms, and concordanced this little corpus using the Wordsmith toolbox.

Using the French corpus

Even this cap-in-hand, unplanned, unbalanced, dubious corpus sheds many insights on problems which had clearly beset us in 1967, and we notice at once several surprising omissions of items which had a high profile in the French corpus. For instance, the ‘3 vt’ section of the *cook* entry (see Figure 1) contains no reference to *laisser cuire*, yet there are a couple of screenfuls of such uses in the corpus (almost all in the imperative: *laissez cuire encore 10 minutes, laissez cuire plus longtemps si...*). Nor does this section mention *cuisiner*, whose transitive use is exemplified in the corpus (*le plat n’est cuisiné ni par nous ni pour nous*). However this transitive *cuisiner* is rare, except as a past participle adjective (*il s’agit aussi bien de plats cuisinés que de desserts*). Similar omissions are also to be found in the entry for *cooking*, where no mention is made of *cuisson*, yet according to corpus evidence it certainly has a strong claim to appear in the ‘1 n’ section (*vérifiez la cuisson, continuez la cuisson, ne change pas de couleur à la cuisson, l’incomparable cuisson au micro-ondes, une cuisson à feu vif, pendant la durée de la cuisson* and so on) and in the ‘2 cpd’ section (in phrases like *eau / jus / sauce / liquide de cuisson* and *mode / méthode / température de cuisson*).

Using parallel corpora

In principle, parallel corpora are interesting for the bilingual lexicographer, and certainly the pairs of sentences extracted for illustrative purposes from the French-English INTERSECT corpus¹⁰ using the Paraconc program¹¹ shed a new light on the hunt for equivalences. However, few commercial dictionary-writing schedules allow time for the lexicographers to browse through data from parallel corpora, at least, not in the simple untreated pair format in which they are shown in Figure 9.

1	a. They take a long time to <i>cook</i> .	b. Ils mettent longtemps à <i>cuire</i> .
2	a. ...delighted by the sweet smell of the bread <i>cooking</i> .	b. ...toute ravie par la tendre odeur du pain en train de <i>cuire</i> .
3	a. When the cake <i>was cooked</i> ...	b. Quand la galette <i>fut cuite</i> ...
4	a. While the cake <i>cooked</i> ...	b. Pendant que la galette <i>cuisait</i> ...
5	a. in order to teach you how to <i>cook</i> –	b. pour vous apprendre à <i>cuisiner</i> .
6	a. She had <i>cooked</i> the ham and eggs	b. Elle avait <i>préparé</i> les œufs et le jambon
7	a. Do not <i>cook</i> or eat in your tent	b. Évitez de <i>préparer</i> et de prendre <i>vos repas</i> à l'intérieur de votre tente;
8	a. Some of the boys could very easily have <i>cooked their meals</i> in...	b. Certains garçons eussent fort bien <i>cuisiné</i> dans...
9	a. he was taking <i>his supper</i> home to Gargan <i>already cooked</i> .	b. <i>Tout prêt le dîner</i> qu'il emportait à Gargan.

Figure 9: Extract from French-English parallel corpora

The principal use of parallel source- and target-language sentences for bilingual lexicographers must be to remind us of equivalents between items in the two languages. Their value of course is that these equivalent pairs occur in the course of text translation, and are not forced into marriage by dictionary compilers. Using parallel corpus data in this form is labour-intensive: each pair has to be read and evaluated. A number of the items in this list are redundant – 1 to 4 inclusive, for instance: no reminder is needed of the *cook* – *cuire* equivalence, either transitively or intransitively. (Information about semantic, syntactic and lexical context of the keywords may more readily be extracted from monolingual corpora.) The problem of redundant material could not be solved by the simple use of 'stop' words, since for instance excluding *cuire* would be too radical (it would certainly remove more than half the results of any search and make impossible any contrastive study of *cuire* and *faire cuire* and *laisser cuire*); and excluding a rarer word like *cuisiner* would lose us the interesting pair 8 as well as the straightforward 5. We need a filter mechanism more open to niceties of grammatical and lexical context, if we are to justify this in terms of the dictionary budget.

From our tiny sample in the list above, this would be hard to do: 1 to 6 inclusive remind us rather unnecessarily of *cuire*, *cuisiner* and *préparer* as equivalents of *cook*. However, the editor of the *cook* entry might be pleased to notice *préparer vos repas* in 7, as might the people compiling *préparer* and *repas*, while 8 is interesting in the context of the *cook* entry,

as well as the entries for *meal* and *cuisiner*. The last pair in 9 illustrate the problems with the use of parallel corpora in commercial lexicography. The phrase *tout prêt le dîner* is too interesting to be summarily dismissed, but too far from any kind of context-free equivalence of *cook* to be of any use to the bilingual lexicographer (who however might spend quite a long expensive moment in reflection before coming to that conclusion). For the moment at least, parallel corpora might contribute more as a resource for the dictionary user rather than the dictionary writer: packaged on a CDROM with an electronic dictionary, they would form a fascinating source of alternative equivalences for the skilled translator.

A smarter program which would tailor the output of the concordancing program to our needs might persuade reference publishers to change their minds about the use of parallel corpora (after all, it took us years to persuade them to let us use corpora at all). However, as far as I know, nothing has been created specifically with dictionary-makers in mind, although there are of course programs which facilitate automatic extraction of candidate translations for machine translation lexicons. Lexicographers working on bilingual dictionaries need some bilingual form of the Word Sketch tool to help them use parallel corpus data within the time constraints of commercial dictionary production.

Looking back on the 1967 entries

Although the early entries seemed to give a reasonable account of *cook* and *cooking*, there are certain points in them that today I find difficult to understand. Why, for instance, had I included *to be head cook and bottle-washer*, already slightly dated, but omitted the more common *too many cooks (spoil the broth)*? Dr. Johnson, on being challenged to explain an error in his magnificent Dictionary, replied "Ignorance, Madam, pure ignorance." "Incompetence" explains most of the bones I am picking with these entries today. The familiar phrase *to do the cooking* is nowhere to be seen. Why is there no sign of the past participle adjective *cooked* – especially since the compounds it forms present distinct translation problems: *cooked food* or *cooked meals* must be rendered as *plats* (= dishes) *cuisinés*, and *cooked breakfast*, not a regular part of the French day, requires a full gloss. As a dictionary user, I mistrust '3 vi' which implies that *faire cuire* and *cuire* are interchangeable, and I have the same problem with '4 vi' over *faire la cuisine* and *cuisiner*. I do not think that by using the *cook* entry anyone would be able to translate *to cook someone a meal* or *to cook breakfast for someone*, or *go and ask cook*, or *cook for 10 minutes in salted water*.

All in all, even before I looked at any lexicographic evidence I was not happy about the 1967 entries, and as I studied the corpus data, and looked at the FrameNet database, it became clear to me that the entries were poor because the source language editor (me) had no clear idea of how the words *cook* and *cooking* actually behaved in natural spoken and written language. I had plucked facts about them from my own native speaker's intuition and other people's dictionaries (one step away from their editors' intuitions); without the support of a massive reading programme there was no way for me, or my French colleague, to marshal all the necessary facts – semantic, syntactic, combinatorial, collocational – about *cook*, *cooking* etc. and their French equivalents. Without such an exhaustive analysis, we could not make a principled selection of the facts most appropriate for the Collins-Robert.

Understanding the English phenomena

What then are the essential points about the source-language headword which must be understood before a bilingual entry is to make sense? FrameNet analyses the verb *cook*, in its literal meanings (ignoring *cooking the books* and *cooking up an excuse*), as belonging to two different frames, provisionally named ‘Apply_heat’ and ‘Cooking_creation’. This fundamental sense distinction escaped me in 1967, and is the key to most of the problems with the entries. It was to be made in relation to the verb *bake* in [Atkins et al. 1988], after a study of corpus data, but that was far in the future. When I was drafting the English entry, it was not clear to me that when you talk about *cooking vegetables* or *cooking fish* this is a different sense from *cooking a meal* or *cooking dinner*. The knock-on effect of overlooking this fact led to the absence of *cuisiner* and *préparer* (used when talking about cooking dishes or meals), of the noun *cuisson* to contrast with *cuisine*, of the past participle *cuisiné* to contrast with *cuit*, and of *cooked meals* and *cooked dinners* and similar phrases.

In 2002, I start writing my entries with the benefit of an analysis of the similarities and differences between these two senses of the verb *cook*:

- Sense 1** LU *cook* in FrameNet ‘Apply_heat’ frame, e.g.
 I cook the onions, the onions are cooking, cook until soft
- Sense 2** LU *cook* in FrameNet ‘Cooking_creation’ frame, e.g.
 he cooked breakfast, he likes cooking.

Sense 1 is a verb with both transitive and intransitive uses, designating the process of causing or undergoing a physical change involving heat, associated with the preparation of food. Sense 2, also a verb with both transitive and intransitive uses, designates the act of creating food, or (of the food itself) being created.

Having established this basic sense division, we must now try to assess in what formal ways the two senses differ. The two lexical units are thoroughly analysed and details recorded in great detail in the FrameNet database, using the type of annotation illustrated in Figure 6. We can therefore draw up a table of similarities and differences according to some of the relevant parameters, as shown in Figure 10.

The two lexical units in Figure 10 correspond to our sense 1 (‘apply heat and change from raw state’) and sense 2 (‘create something else by doing that’).

		LU 1 Apply_heat	LU 2 Creation
1	Object of transitive	raw food	prepared food
2	Subject of transitive	a) person b) oven <i>etc.</i>	person
3	Subject of intransitive	food	person
4	'activity/achievement' interpretation	atelic	telic
5	Causative/inchoative alternation	yes	no
6	Benefactive indirect object	no	yes
7	Benefactive PP-for	no	yes
8	Omitted direct object interpreted as specific food already mentioned in discourse	yes	no
9	Omitted direct object interpreted as indefinite 'food'	no	yes

Figure 10: Similarities and differences in the two senses of verb *cook*

Rows 1-3 in the table compare the semantic types of nouns which fill the subject and direct object slots of the LUs. As row 1 shows, food itself functions as the direct object in both senses; however, it would be more accurate to say that with sense 1 (*cook the peas in the usual way*) the food is something raw about to be processed, while with sense 2 (*I'm cooking a curry*) it is something which has been prepared by processing the raw foodstuffs. **Row 2** records that with sense 1 we can find the source of the heat (FrameNet's frame element 'Heating_Instrument') in subject position (*my new oven cooks meat really well*), as well as in a PP.comp (*I prefer to cook it in the microwave*), while this is not possible with sense 2 (**my new oven cooked me breakfast*).

Row 4: in considering senses 1 and 2, we notice that the object of *cook* in sense 1 is what is being processed, while the object of sense 2 is what is being produced by the process (and served and eaten). Our instinct is to distinguish these senses on the basis of the type of nouns filling the object slot – raw food (*meat, carrots, onions*) for sense 1, and prepared food (*pie, breakfast, meal*) for sense 2. For FrameNet, however, the distinction between sense 1 and sense 2 shown in row 4 does not inherently depend on the semantics of the noun filling the direct object slot: it depends on whether the event is construed as an activity (atelic) or an accomplishment (telic)¹². Thus, *cook the minced meat until browned* is clearly apply_heat (sense 1), while *she cooked Sunday lunch in an hour* is clearly Cooking_creation (sense 2). Both of these examples may be rephrased with the same noun as direct object: *cook the potatoes until browned* (sense 1) and *she cooked the potatoes in an hour* (sense 2). The distinction depends on our knowledge that with the verb in sense 1, the potatoes are not the finished product, while with sense 2 they are what is produced by the process (and served up). For the sake of the poor dictionary users, however, the eventual bilingual entry must select quite distinct nouns to exemplify this difference: perhaps *meat, vegetables* for sense 1 and *meal, dish* for sense 2.

Row 5 highlights the fact that sense 1 (*while she was cooking the pasta / while the pasta was cooking*) participates in the causative-inchoative alternation [Levin 1993], discussed in relation to the verb *bake* in [Atkins et al. 1988]: the object of the transitive use in sense 1 becomes the subject of the intransitive, while in sense 2 it does not. In sentences like *breakfast was cooking*, we have to interpret *cooking* as ‘being processed’ (sense 1), not ‘being produced’ (sense 2).

In **rows 6 and 7**, we note that only sense 2 allows the person for whom the action is done (in FrameNet terminology, the frame element RECIPIENT) to be expressed both as the indirect object (*I cooked him a meal*) and as the prepositional object in the PP.comp (*I cooked a meal for him*).

Rows 8 and 9 focus on an problem which I found impossible to solve when I was writing the 1967 entries. I could think up sentences like *cook for 20 minutes or until browned*, and *she has been cooking all day*, and indeed could recognise that in some obscure way the direct object was unexpressed in both cases, but despite the obvious differences in the target-language, I could not make a clear distinction in the meanings involved, and indeed omitted the *cook for 20 minutes* construction altogether, as the entry in Figure 1 shows. I probably could not decide where to put it – it would have looked uncomfortable grammatically in ‘3 vi’ and ill at ease semantically in ‘4 vi’. It is for problems like this that every lexicographer needs a linguist on hand. Thanks to the sophisticated method which FrameNet has of distinguishing such constructions, this is no longer a problem. Sentences like *cook for 20 minutes* (an instance of the ‘instructional imperative’ discussed in [Atkins et al. 1988]) occur only in recipes. The omitted direct object is the specific foodstuff being cooked, which is known to the recipe reader. In FrameNet terms, this is treated as a case of Constructional Null Instantiation (CNI), where the grammatical construction licenses the omission of the (known) direct object. The situation is different in the case of *she has been cooking all day*, where the omitted direct object does not need to be specifically known to either the speaker or hearer for the sentence to be understood naturally. FrameNet terms this type of object omission Indefinite Null Instantiation (INI). As rows 8 and 9 show, both sense 1 and sense 2 allow null instantiation of the object, but in the first this is CNI and in the second INI. Our dictionary entry must make this clear, irrespective of whether these phenomena are mirrored in the target language.

The equivalences in the entries

Mention has already been made of some target-language omissions in the 1967 entries which became apparent as soon as a French corpus was available: *laisser cuire* and *cuisiner* from the *cook* entry and *cuisson* from the *cooking* entry. There were however many other problems that had to be discussed with my francophone collaborator, Thierry Fontenelle¹³ before the bilingual entry could be fully drafted.

Starting from two senses of *cook* derived from the FrameNet analysis, we determined that sense 1 (change from raw state) would require a French equivalent using the verb *cuire*, while sense 2 (produce, create) would be translated by a more general verb like *préparer* (prepare) or even *faire* (make). Unlike the earlier version, the new entry will be quite explicit

about these two senses, since I believe that an understanding of this will help the English-speaking user to appreciate why *cuire* (everyone's first choice as the French equivalent of *cook*) does not fit all contexts.

Having decided on the *cuire* – *préparer* distinction, we began our discussion of the real problems of equivalence. I had assembled FrameNet's account of the two senses of *cook*, together with the concordances from the French corpus for *cuire*, *cuisine*, *cuisinier*, and *cuisson* and their inflected forms. I had inserted sketchy English equivalents inserted against some of the examples, added notes explaining the difficulties I envisaged for the anglophone user of these entries, and listed my principal queries for Fontenelle. I shall now summarize our discussions of each of these, using 'sense 1' to designate the Apply_heat sense, and sense 2 the Cooking_creation sense; this distinction was of course not made in the original entries. The revised draft of part of the entry for *cook* is shown in Figure 11.

<p>cook I verb 1 (change from raw state) 1a vt [<i>person</i>] <i>food</i> faire cuire, cuire; [<i>oven</i>] <i>food</i> cuire. 1b vi [<i>food</i>] cuire. the soup cooking on the fire la soupe qui cuisait au feu. 1c vi (<i>in recipes</i>) (faire) cuire, (+ <i>time expression</i>) laisser cuire. cook in a hot oven (faites) cuire à four très chaud; cook for 10 minutes</p>	<p>(<i>stressing actual cooking</i>) cuisiner. to cook sb a meal, to cook a meal for sb préparer un repas à qn. 2b vi [<i>person</i>] (<i>generally</i>) faire la cuisine; (<i>stressing actual cooking</i>) cuisiner. it's Paul who cooks c'est Paul qui fait la cuisine; to cook outside in the open cuisiner en plein air; she cooks well elle fait bien la cuisine, elle cuisine bien; (<i>in lists</i>) I can't sew or cook je ne sais ni coudre ni cuisiner; she used to cook for them elle leur préparait des repas.</p>
---	---

Figure 11: Revised draft for the verb entry

1. *cook* sense 1a vt : *cuire* / *faire cuire* / *laisser cuire* / *cuisiner*

This is a complex query, dealing as it does with four possible equivalences of *cook* in sentences like *I prefer to cook it in the microwave* and *cook the pasta in salted water for 10-12 minutes*. Setting aside for the moment the knotty problem of distinguishing between *cuire* and *faire cuire*, we turned first to *laisser cuire*, and noticed at once that in the many sentences where this expression was to be found, no direct object was ever expressed, although the verb phrase is transitive. We decided to pair this up with the instructional imperative (see below), since, like *cook* in this use, the omitted direct object was constructionally licensed. Here was another case of CNI.

We then looked at the concordances in which the verb *cuisiner* was used transitively: only seven out of more than 100 instances. When we checked the nouns¹⁴ occurring in the direct object slot, we realised that *cuisiner* was in fact the equivalent of sense 2 (produce, create) and not of sense 1 (process from raw state).

We set about the task of discovering and transmitting to the dictionary user the difference (if any) between *cuire* in its transitive use, and *faire cuire*. We extracted a large number of facts from the corpus, spent a disproportionate amount of time discussing this problem, but made virtually no progress. This is frequently the case in lexicography: you spend hours puzzling something out, often discussing it with colleagues. Sometimes you come to understand a lot more about the behaviour of your headword, sometimes it remains opaque. When you do find yourself able to articulate a useful distinction, you then try to improve the entry, spend far too long inserting, deleting, and amending, and finally decide that it must be left as it is – not perfect, but the best that can be done given the space constraints.

In our study, we looked first at the semantic types of the nouns which occur in the direct object slot in corpus sentences, making lists¹⁵ of most of the noun phrases found in that slot. We concluded tentatively that *cuire* seems to favour wheat-based products, but not exclusively, and *faire cuire* seems to favour vegetables, meat, fish, seafood, eggs, game... but again not exclusively. We considered the idea that substances such as bread or dough might be intuitively believed to cook in a different way from other substances such as fish, meat and eggs, so that *cuire* was more acceptable with the first group than with the second, which preferred *faire cuire*. We could find no solid evidence for this.

Turning next to grammatical context, we considered the types of noun phrases which occurred in the direct object slot with *cuire* and with *faire cuire* respectively, and in particular the types of determiner, but found no systematicity here either. We checked the use of these expressions in the past participle, but found nothing there either. Certain syntactic contexts seem to prefer *cuire* over *faire cuire* – manner adverbials, modal verbs, and infinitive constructions of purpose – but this might come about simply because *cuire* is less of a mouthful than *faire cuire*. Only in one aspect of their behaviour, the use in the imperative mood, was there any clear difference: we found no instance of *cuisez*, but many of *faites cuire*. However, we were forced to admit that we could not find any systematic differences, or even identify any consistent selectional preferences, in the use of these two expressions. In every concordance line we looked at, when the verb was transitive and used in the meaning of ‘change state from raw to cooked’, *cuire* was apparently substitutable for *faire cuire*, and vice versa.

3 vt (a) food (faire) cuire

Figure 12: part of original entry for *cook*

We came to the conclusion that the 1967 entry, shown in Figure 11, though brief, had been correct. The bracketing of *faire*, according to the Style Guide, indicates that this is an optional element in the translation. Remembering, however, that on returning to the entry 35 years later I had been unable to trust it, I determined to expand this part a little, in order to reassure other dictionary users.

2. *cook* sense 1b vi : *cuire*

The concordances made it clear that *cuire* is the principal, and the safest, equivalent of *cook* in sentences like *a stew was put to cook in a saucepan*. There was no need for any further

information in this section, but in order to distinguish it from the ‘intransitive’ use in the following section, we included a short example.

3. *cook sense 1c vi : instructional imperative*

The instructional imperative, superficially intransitive but essentially transitive with an unexpressed definite object, requires some thought: how to classify it, how much of the grammar to explain to anglophone or francophone users, and how to translate it. Following the Style Guide, we called it an intransitive verb. Like English, French has a similar pseudo-intransitive construction, so there was no need to explain the grammar to either the English or the French user. It was obvious from the evidence of the concordances that this was the place for *laisser cuire*. This translates as ‘leave to cook’, an unexceptionable phrase but one which does not occur much in the English corpus data we had.

In our French corpus of some 2,000 sentences, *laisser cuire* was very prominent, and occurred exclusively in sentences like those in Figure 13:

Laissez	cuire	5 minutes, réservez.
Laissez	cuire	petite ébullition pendant 2 heures.
Laissez	cuire	encore 10 minutes, passez au chinois.
Laissez	cuire	doucement 10 minutes...
Laissez	cuire	un petit peu jusqu’à ce que...

Figure 13 : Concordance lines for *laisser cuire*

Every one of the *laisser cuire* lines occurred in the imperative (*laissez cuire*); every one came from a recipe¹⁶; none of them had an expressed direct object, and every one of them included a duration expression (‘5 minutes’, ‘until...’ etc.).

However, we had many corpus examples, also from recipes, of *faites cuire* and of the bare infinitive *cuire* used as an imperative. As (*faire*) *cuire* is always a safe option, it was tempting to omit *laisser cuire*, which seemed to require a time expression, but the frequency of the latter phrase in the French corpus and its nigh-perfect match with the English recipe-ese usage earned it a place in the entry. However, unless we were sure (and we weren’t) that the three phrases were interchangeable, we had to puzzle out what the differences were, and guide the user to the correct option.

A study of the corpus evidence leads to the conclusion that the difference between (*faire*) *cuire* on the one hand, and *laisser cuire* on the other is the following. *Faites cuire* and *cuire* occur in texts at the point in the instructions where the actual cooking process begins. *Laisser cuire* occurs where some previous instruction has initiated the cooking process, e.g.

*vérifiez la cuisson, laissez cuire plus longtemps si ...
versez dans la sauce, laissez cuire encore 10 minutes*¹⁷

It proved impossible to condense that information intelligibly, and we had to content ourselves with prefacing *laisser cuire* with ‘+ time expression’, which was the truth, if not the whole truth.

4. cook sense 2a vt : translation, including benefactives

Our starting point in discussing the translations of *cook* was to make a systematic distinction between sense 1 (*cuire* etc.) and sense 2 (*préparer, faire*). Neither *préparer* nor *faire* appears in the 1967 entry, which indeed addresses only what we are calling sense 1 of the English verb. A study of the French corpus data and other material¹⁸ left us in no doubt that *préparer* and *faire* must be offered as direct equivalents.

She would	cook	them all a good breakfast
Toma	cooked	me a man-sized breakfast
I came home to	cook	myself lunch
...to come home and	cook	three meals a week for her family
I'm	cooking	vegetable curry for them

Figure 14: concordance lines showing benefactives

Figure 14 shows instances of the element RECIPIENT in the cooking_creation frame (the person for whom the action is performed) expressed in two ways: as the indirect object of *cook* (*cook them all a good breakfast*) and within a prepositional phrase as the object of *for* (*cooking vegetable curry for them*). These constructions are part of the valence of our headword and it is therefore essential that anglophone dictionary users find in the entry enough information to allow them to produce the correct French equivalent; we added to 'vt' an example phrase in the two formulations, see Figure 11.

5. cook sense 2b vi : faire la cuisine / cuisiner

There are a number of different points to be taken into consideration when producing the translations for this section of the entry. We know from the corpus that *cuisiner* (but not *faire la cuisine*) functions like *cook* sense 2 in two significant respects. It has transitive uses (*le plat n'est cuisiné ni par nous ni pour nous*), even if these are much less salient than the transitive uses of *cook*. It also supports null instantiation of its indefinite object (INI) : the corpus offers many superficially intransitive uses where the unexpressed object is indefinite (*la fille savait pas cuisiner*).

It is tempting to leave *cuisiner* as the only direct translation here, but *faire la cuisine* – literally 'do the cooking' – is simply too common to omit¹⁹. After comparing usages in the French corpus, Fontenelle came to the conclusion that on the whole these verb phrases are freely interchangeable, except that *cuisiner* focusses more on the actual cooking process, while *faire la cuisine* carries more the general idea of being responsible for preparing food. In particular, as the 'and/or' list in the Word Sketch shown in Figure 3 indicates, *cook* often occurs in lists of activities, where *faire la cuisine* would be unlikely in French²⁰. We added a couple of examples to try to indicate this distinction.

The FrameNet RECIPIENT can also be expressed with this intransitive use of *cook*, not however as an indirect object (which occurs only with transitives – see point 5) but in a prepositional phrase headed by *for* (*she was told to cook for the soldiers, I cook for you and I*

keep house for you): we add an example to show this in the entry, thus demonstrating that it is impossible to improve a dictionary without adding to its length.

6. noun *cook* as name and form of address

Before we leave *cook*, we have to address an omission in ‘1 n’ in the 1967 entry. The inclusion of the first example (*she is a good cook*) immediately after the direct translation (*cuisinier m, ière f*) implicitly distinguishes between the use of the noun as designating a member of a trade or profession (*he was a hotel cook, the cook and the gardener*), and the deverbial noun (*she’s a good cook*), but it gives no help with the French equivalent of the ‘bare’ noun (*I must speak to cook*). This usage is pointed up in a column in the Word Sketch screen not visible in Figure 3. Nor does the 1967 entry include the use of *cook* as a form of address (*Good morning, Cook*). This turns out to pose a knotty problem of translation, and to do it justice in a dictionary would involve several more examples,²¹ together with explanations for the English user. Manifestly, in 2002, to cover this point in enough detail would clearly demand a disproportionate amount of space in a general dictionary (as opposed to one for translators of historical novels). Figure 15 shows what the entry for the noun *cook* would have to look like if this point is to be covered, even inadequately.

<p>cook noun 1 : to be a good etc. cook cuisinier bien, bien faire la cuisine; she’s a splendid cook elle cuisine merveilleusement bien. 2 (in home, hotel etc.) cuisinier m, -ière f. (used as a name) I must speak to Cook je dois parler à la</p>
--

Figure 15: Draft entry for noun *cook*

7. cooked ptp adj : cuit / cuisiné

The past participle adjective *cooked* does not figure in the original entry, but since the only translation given for the transitive verb is (*faire cuire*), the user must infer that the equivalent of *cooked* is *cuit*. I probably thought so at the time that I wrote the entry. Figure 16 shows some corpus lines which seem to confirm that impression.

il est comestible	cuit,	mais vénéneux cru.
Il est consommé cru ou	cuit.	
La crème est	cuite	lorsqu'elle nappe la cuiller.
viande de cheval crue ou insuffisamment	cuite	
Ils empêchent les aliments	cuits	de rancir.

Figure 16 : Some corpus instances of *cuit*²²

However, I also find in the French corpus the use of the past participle *cuisiné* as an adjective, particularly in the phrase *plats cuisinés* ('cooked food'). It seems that the past participles *cuit* and *cuisiné* reflect the sense distinctions already established between sense 1 of the verb (change from raw state) and sense 2 (prepare, produce), and this must be indicated in the entry for the past participle adjective, a draft of which is shown in Figure 17.

<p>cooked ptp adj 1 (not raw) food <i>cuit</i>. cooked fruit / vegetables légumes / fruits <i>cuits</i>; cooked ham jambon <i>cuit</i>. 2 (prepared) meal, dish <i>cuisiné</i>. cooked meals, cooked food plats <i>cuisinés</i>; cooked breakfast petit déjeuner complet à l'anglaise; cooked dinners (<i>as opp. to sandwiches</i>) repas <i>chauds</i></p>

Figure 17 : Draft entry for *cooked*

8. *cooking n : cuisine / cuisson*

In the framework for *cooking*, based on evidence in the English corpus, the noun has at least three senses:

1. the process of changing something from its raw state (*cooking method*);
2. the activity of producing food (*they share the cooking*);
3. the food produced (*his mother's cooking*).

Sense 1, the least frequent, is related to the Apply_heat sense of the verb; senses 2 and 3 are related to the Cooking_creation sense of the verb. According to the randomly selected sample of the BNC, sense 3 is by far the most salient. The candidates for inclusion as translations are the nouns *cuisine* and *cuisson*. We noted that the original entry contained no reference to *cuisson*, yet *cooking* is the only real English equivalent of this word, although in context it may be translated in many different ways, as the table in Figure 18 shows.

<i>Vérifiez la cuisson, laissez cuire...</i>	check to see if it is cooked ...
Préparation : 30 mn. <i>Cuisson</i> : 15 mn.	cooking time 15 minutes
<i>...ne change pas de couleur à la cuisson</i>	...does not change colour when cooked
<i>l'incomparable cuisson au micro-ondes</i>	...microwave cooking
<i>...ail, persil, thym et laurier avec un concassé de tomates dans une cuisson au vin blanc</i>	...cooked together in white wine
<i>des fours à cuisson automatique</i>	automatic ovens
<i>après une heure de cuisson à feu très doux</i>	...after simmering for an hour
<i>La plancha permet une cuisson à feu vif sans matières grasses</i>	in the Plancha you can cook on high heat without fat
<i>le temps d'une mayonnaise ou de la cuisson d'un oeuf dur</i>	in the time it takes to... hard-boil an egg
<i>les consignes concernant la cuisson d'un plat spécifique</i>	advice about ways of cooking a specific dish
<i>Le micro-ondes est parfait pour la cuisson des carottes</i>	...perfect for cooking carrots ...

Figure 18 : Some English equivalents of *cuisson*

The material in Figure 18 makes us wary of giving *cuisson* without explanation as a straightforward equivalent of the noun *cooking*, and it will certainly pose serious problems for the lexicographer compiling the *cuisson* entry. We obviously cannot ignore it. Fontenelle points out that the default translation (always a priority for bilingual lexicographers) of the noun *cooking* is *cuisine*, which would translate the vast majority of the corpus instances of this word, such as those shown in Figure 19.

the loch provided water for	cooking	and drinking .
Natural gas is used for	cooking	
I hate the	cooking	and the cleaning
Molly's superb	cooking	
Microwave	cooking	is now back in vogue
he's an authority on Chinese	cooking	
The English use it in their	cooking	
her plain	cooking	is quite good
a taste for wine and continental	cooking	
the scents of north African	cooking	

Figure 19: Some concordances of the noun *cooking*

The difference in meaning and use between *cuisine* in its cooking sense (as opposed to 'kitchen') and *cuisson* were not initially apparent. The less familiar noun *cuisson* designates a process – it is of course the nominalisation of *cuire*, which corresponds to the change-from-raw-state sense of *cook*. *Cuisson*, but not *cuisine*, occurs with duration phrases (*la cuisson doit durer 5 minutes* 'it should cook for 5 minutes', *pendant la cuisson* 'during the cooking'), and is used to refer to specific cooking conditions, as in *cuisson rapide* ('fast cooking'). Historically, *cuisine* gave rise to *cuisiner* – explaining its sense of creating food,

see points 4 and 7 above; it refers to the general activity relating to the preparation of meals (*la cuisine au gaz* ‘gas cooking’, *la cuisine au micro-ondes* ‘microwave cooking’).²³ Our discussions end in the draft entry for *cooking* shown in Figure 20.

<p>cooking I noun 1 (see <i>vb 2: activity, food</i>) <i>cuisine f. to do the cooking faire la cuisine; French cooking la cuisine française.</i> 2 (see <i>vb 1: process</i>) <i>cuisson f. slow cooking cuisson lente.</i> II modifier 1 <i>de cuisine. cooking smell/ ustensil odeur f/ ustensile m de cuisine.</i> 2 <i>de cuisson. cooking liquid / method liquide m / méthode f de cuisson.</i> 3 <i>apples, chocolate etc. à cuire.</i></p>
--

Figure 20 : Draft entry for *cooking*

When it comes to writing the dictionary entry, an accumulation of facts makes us change the sense division and ordering from the three senses in the framework, set out at the start of this section 8. First, the order of senses in the bilingual entry should if possible reflect frequency in the English corpus²⁴, which means putting ‘food produced’ as the first sense in the entry. Second, *cuisine* is far commoner than *cuisson* in the French corpus, which suggests that it is the ‘safest’ word for the English speaker to choose as a translation of *cooking*. Third, comparison of concordances for the English and French words makes it clear that the framework senses 2 and 3 of *cooking* should be translated by *cuisine*, and sense 1 by *cuisson*. Finally, since senses 2 and 3 have the same French equivalent, we can save space by collapsing them in the entry.²⁵

Conclusion

In this paper I have tried to set out in detail the transfiguration wrought in practical lexicography by the advent of the computer. Writing dictionaries in 1967, even with the help of a citation bank on index cards, depended mainly on introspection and on discussion with one’s colleagues and other informants – introspection again, at one remove. As [Hanks 2000] points out, the entries that we wrote then reflected what he calls ‘cognitive salience’ – things that stand out in our minds when we think about language. Corpus lexicography has taught us that packing an entry with cognitively salient items does not produce a good description of naturally occurring language. We have no access through introspection to how the language really behaves out there in the linguistic community, or even to how we ourselves use language in speaking and writing. Idioms like *cook someone’s goose* are cognitively salient, but extremely infrequent in corpus data. Expressions like *cook him breakfast* occur so frequently that they are not cognitively salient and are often overlooked in the dictionaries where they should occur, such as those for language learners.

But wealth of data alone does not make a good dictionary.²⁶ It simply swamps us with a dazzling array of facts and no systematic way of evaluating them. The craft of lexicography demands not only the ability to collect data, but also the ability to make sense of it. We have to be able to see the whole picture of how our headword behaves in natural language before we can decide what information is most needed by the people who will use the dictionary we are writing. Then we need to set out these facts in an intelligible and orderly way. Without a sound theoretical basis we cannot carry out these tasks successfully. The most significant difference, I believe, between the 1967 lexicography and that of today is that in the interval my approach to lexicography has benefited from the insights of linguistics.

In 2002, we have plenty of evidence about what Hanks calls ‘socially salient’ facts. A corpus like the BNC offers us literally thousands of citations for the core vocabulary of the language. Tools like the Word Sketch program, and resources like the FrameNet database, extract socially salient facts from this mass of data. Linguistic theory, particularly recent work in lexical semantics, can light the way to better lexicography. At last we are in a position to begin to reflect performance, and not our own competence, in our 21st century dictionary entries.

Acknowledgements

I’m glad to have this opportunity of thanking my CREFD colleagues for many years of friendship and collaboration on this team-built dictionary: Richard Thomas, Collins’ editorial director of the project; my French co-authors of the English-French entries, Marie-Noëlle Lamy, Hélène Lewis and Renée Birks, and our counterparts on the French-English side, Alain Duval, Rosemary Milne, Pierre-Henri Cousin and Lorna Sinclair Knight. Michel Poté’s prototype discussions on the letter H in 1967-8 were invaluable, as later were those with Geneviève McMillan. My warmest thanks go to Thierry Fontenelle, who acted as my francophone colleague for the ‘Now’ lexicography, for his knowledgeable and careful contribution to, and patience in, our electronic discussions on detailed points of equivalence. I am grateful to my colleagues on the FrameNet team at ICSI, Berkeley, under the leadership of Charles Fillmore, for their exhaustive analysis of *cook*. Many thanks, too, to everyone who helped me collect citations (in the absence of a French National Corpus) for *cuire* and related words: Kate Beeching, Pierrette Bouillon, Michael Clark, Marie-Hélène Corréard, Alain Duval, Thierry Fontenelle, Adam Kilgarriff, Raf Salkie, Penny Silva, Natalie Pomier, and Michael Rundell. Thierry Fontenelle, Valerie Grundy, Adam Kilgarriff, Miriam Petruck, Michael Rundell and Raf Salkie all made valuable comments on earlier versions of this paper. Finally, the ‘Now’ lexicography owes much to what I have learned over the past 15 years or so from many colleagues I’ve worked with, in particular two linguists, Charles Fillmore and Beth Levin, and two lexicographers, Patrick Hanks and Michael Rundell. Thank you all.

Endnotes

¹ The dictionaries I used as ‘lexicographic evidence’ were an early edition of the Concise Oxford English Dictionary, the Oxford Advanced Learner’s Dictionary (1963), and the one-volume Harrap English and French Dictionary; later were added the Random House College Dictionary 2nd edition, and eventually the *Petit Robert*.

² The lexeme *cook* unites the word-forms *cook*, *cooks*, *cooking* and *cooked*, and may be analysed into lexical units; we may think of a lexeme as corresponding to a dictionary headword, and a lexical unit to a dictionary sense.

³ The *target language* is, for instance, French in an English-into-French bilingual dictionary entry, and the *source language* in such a dictionary is English.

⁴ See <http://info.ox.ac.uk/bnc>

⁵ WordSmith, by Mike Scott, is a useful, inexpensive corpus-querying utility; details of the program are available at <http://www.liv.ac.uk/~ms2928/homepage.html>. Dictionary publishers with large corpus resources tend to use more powerful concordancing software.

⁶ See <http://wasps.itri.bton.ac.uk/> for demo version.

⁷ This research project, of considerable importance to professional lexicographers, is based in the International Computer Science Institute, Berkeley, California, and led by Charles J. Fillmore, whose work in frame semantics and construction grammar informs the lexicography. A full account of this project will be given by the FrameNet team in a forthcoming edition of the *International Journal of Lexicography*, to be guest-edited by Thierry Fontenelle and scheduled for December 2002.

⁸ See <http://www.icsi.berkeley.edu/~framenet/>

⁹ As the entry is not complete, the frame names are not yet finalized.

¹⁰ See <http://www.brighton.ac.uk/edusport/languages/html/intersect.html>

¹¹ Available from Athelstan. Information at <http://www.athel.com/>

¹² I owe this insight to Charles Fillmore.

¹³ Thierry Fontenelle worked with me on the target-language problems discussed here. His contribution was all the more valuable because of his intimate knowledge of both the dictionary and FrameNet lexicography. His conversion of the Collins-Robert English-French Dictionary into a Mel'čukian-type electronic database is described in [Fontenelle 1997], and his application of this database in a FrameNet perspective is discussed in [Fontenelle 2000].

¹⁴ These nouns included *d'autres denrées* (other foods), *des meringues* (meringues), *des spaghettis* (spaghetti), *les tartes à la crème* (custard tarts) and *les plats chinois* (Chinese dishes).

¹⁵ With *cuire* the list included *son pain*, *le manioc*, *du pâté sablé*, *nos aliments*, *une pâte surgelée*, *des baguettes*, *du pain*, *des pâtons surgelés*, *le kouglof*, *le pain*, *un mets*, *les aliments*, *les saucisses*, *un poulet*, *deux repas quotidiens*, *le blé*, *confitures et gelées*, *les cocos*, *le pigeon*, *cabillaud*, *pain de campagne*, *le magret de canard*, *plat de côtes*, *le morceau de lotte*, *la crème*... In English: his bread, manioc, shortcrust pastry, our foodstuffs, frozen pastry, loaves, some bread, frozen dough sticks, kugelhof, bread, a dish, foodstuffs, sausages, a chicken, two meals a day, wheat, jams and jellies, coconuts, pigeon, cod, some farmhouse bread, some breast of duck, dish of ribs, the piece of monkfish, cream.

With *faire cuire* we found *pommes de terre*, *l'agneau*, *toutes les espèces de poules*, *une viande*, *un poisson*, *les oeufs*, *leur petite soupe*, *ces deux légumes*, *des châtaignes*, *un oeuf*, *mon omelette*, *les fruits*, *mes champignons*, *du riz*, *des graines de blé*, *les artichauts*, *des pommes*, *le sucre*, *les filets de dorade*, *les canards*, *les pâtes*, *les tourteaux*, *la pâte*, *les petits oignons*, *les asperges*, *carottes et poireaux*, *les langoustines*, *les côtes d'agneau*, *leur frichti*... In English: potatoes, the lamb, all kinds of chicken, a piece of meat, a fish, eggs, their soup, these two vegetables, chestnuts, an egg, my omelette, fruit, my mushrooms, rice, wheat grains, artichokes, apples, sugar, bream fillets, ducks, pasta, crabs, pastry, little onions, asparagus, carrots and leeks, scampi, lamb chops, their grub...

¹⁶ The larger context available to us made us aware that in many cases an English recipe would not have used *cook* at all, but a more specific hyponym such as *simmer*, *fry*, *boil*, *bake* and so on – something to note about *laisser cuire* for the *cuire* entry, but not something to try to include in our *cook* entry.

¹⁷ In English: check and cook longer if ..., pour into the sauce and cook for 10 minutes more.

¹⁸ In researching this aspect of the translation, Fontenelle noted: “I was able to find more than 16,000 occurrences of *préparer* + *petit déjeuner* on Google. **Cuisiner* + *petit déjeuner* would not be found at all and *faire* + *petit déjeuner* would be found but much much less frequently than *préparer*” – a comment which in itself more than justifies corpus use in lexicography.

¹⁹ The French *cuisiner* and *faire la cuisine* almost exactly parallel in semantic nuance the English *cook* and *do the cooking*, the only difference being the relative salience of *faire la cuisine* in the French corpus and comparative infrequency of *do the cooking* in the English corpus. In a dictionary for encoding users, however, frequency matters (we are all looking for the most natural-sounding expression), and our entry must address this problem.

²⁰ For the same reason – related to analogy – that we would be reluctant to substitute *do the cooking* for *cook* in a list like *he can clean and cook and sew*.

²¹ For *I must speak to cook* (or *Cook*), we would have to show options on the French translation depending on context. *Je dois parler à la cuisinière* (assuming a female cook) is what the employer – owner perhaps of a stately home or a private hotel – would say in general conversation, but if talking to another employee they are more likely to say *Je dois parler à Madame Duval*, or whatever the lady’s name is. In the case of *Good morning, Cook* the French would have to be *Bonjour, Madame Duval*. *Madame* alone is too formal for an employer to use to someone working for her, and unlike *docteur etc.* the noun *cuisinier* cannot be used in vocatives.

²² In English: *it is edible cooked, but poisonous raw; it is eaten raw or cooked; the cream is cooked when it coats the spoon; horse meat raw or insufficiently cooked; they prevent cooked food from going bad.*

²³ We find two disconcerting examples relating to the familiar phrase *microwave cooking*:

- (a) La réfrigération et *la cuisine au micro-ondes* ont eu des répercussions importantes sur l’emballage... ‘Refrigeration and microwave cooking have had serious repercussions on packaging ...’
- (b) Assaisonner le lièvre avec sel Vital: essayez aussi l’incomparable *cuisson au micro-ondes* (15 mn pour 500 g)... ‘Season the hare with Vital salt: sample the incomparable microwave cooking...’

In (a) the topic is the change brought about in packaging by the development of refrigeration and microwave cooking – far removed from the heat and smells of the kitchen: here we have *la cuisine au micro-ondes*. The citation in (b) is from a recipe, instructions on cooking hare, and for this context *la cuisson au micro-ondes* is more appropriate.

²⁴ It is always better if possible to reflect source-language frequency in a bilingual dictionary: the commonest use probably generates the largest number of look-ups. However, the order of senses in the verb entry reflected the concept of ‘basic’ or ‘core’ sense (*Apply_heat*) from which a more complex sense (*Cooking_creation*) is derived. This logical approach makes it easier to guide the encoding user through the intricacies of the foreign language equivalents. The decision to follow logical order in the verb entry and frequency order in the noun entry rests on the need to communicate as efficiently as possible with the dictionary user. Lack of correspondence between entries for morphologically related verbs and nouns may disturb computers but rarely worries human users.

²⁵ This is a good example of why the ‘senses’ in a bilingual dictionary rarely reflect accurately the semantics of the source-language word.

²⁶ Michael Rundell pointed this out to me, having benefited himself from the contribution of linguistic theory to practical lexicography in the new Macmillan learners’ dictionary [Rundell 2002].

References

- [Atkins 1993] Atkins, B. T. S., 1993. Theoretical Lexicography and its Relation to Dictionary-Making, in: W. Frawley (ed.) *Dictionaries: the Journal of the Dictionary Society of North America*. DSN, Cleveland, Ohio, USA. 4-43.
- [Atkins et al. 1988] Atkins, B. T. S., J. Kegl & B. Levin, 1998. Anatomy of a Verb Entry: from Linguistic Theory to Lexicographic Practice, in: *International Journal of Lexicography*, 1:2, Oxford University Press, Oxford. 84-126. Also in A. Zampolli, N. Calzolari & M. Palmer (eds.) *Current Issues in Computational Linguistics: In Honour of Don Walker, Linguistica Computazionale IX-X*. (1994). Giardini Editori, Pisa, Italy. 237-266.
- [Fillmore 1995] Fillmore, Charles J., 1995. The Hard Road From Verbs To Nouns, in: M. Chen and O. Tzeng (eds.) *In Honour of William S-Y Wang: Interdisciplinary Studies on Language and Language Change*. Pyramid Press 105-129.
- [Fontenelle 1997] Fontenelle, Thierry, 1997. *Turning a bilingual dictionary into a lexical-semantic database*, Lexicographica Series Maior 79, Max Niemeyer Verlag, Tübingen, Germany.
- [Fontenelle 2000] Fontenelle, Thierry, 2000. A Bilingual Lexical Database for Frame Semantics, *International Journal of Lexicography*, Oxford University Press, 13:4, 232-248.
- [Hanks 2000] Hanks, P., 2000. Contributions of Lexicography and Corpus Linguistics to a Theory of Language Performance, in: U. Heid et al. (eds.) *Proceedings of Ninth Euralex International Congress*, IMS Stuttgart University, Stuttgart, Germany, 1:3-13.
- [Kilgarriff & Tugwell 2001] Kilgarriff, A. & D. Tugwell, 2001. Word Sketch: Extraction and Display of Significant Collocations for Lexicography, ACL Workshop on Collocation, Toulouse, July 7 2001; available on line from <http://wasps.itri.bton.ac.uk/>, or as ITRI technical report ITRI-01-12 from <http://www.itri.bton.ac.uk/techindex.html>.
- [Levin 1993] Levin, Beth 1993. *English Verb Classes and Alternations: A Preliminary Investigation*, University of Chicago Press, Chicago, USA.
- [Rundell 2002] Rundell, Michael 2002. (Editor-in-Chief) *Macmillan English Dictionary for Advanced Learners*, Macmillan Publishers Ltd., Oxford, 2002.