

What are concordances for? Getting multimodal concordances to perform neat tricks in the university teaching and testing cycle

ANTHONY BALDRY

1. INTRODUCTION

Developments in multimodal studies of meaning making have given rise in recent years to descriptions and teaching materials relating to printed pages, web pages, film texts and genres (O'Halloran, 2004; Baldry, Thibault, 2006a) in fields as disparate as translation and subtitling (Taylor, 2004), film studies (Ackerley, Coccetta, *in press*, Dalziel, Metelli, *in press*) and medicine (Baldry, Guardamagna, *in press*). But research into testing activities relating to these genres, and multimodal genres in general, has lagged behind. In the belief that the integration of multimodal modules into university syllabuses cannot occur properly without innovative text-based solutions in the teaching-testing cycle, this paper explores multimodal concordancing as an area of research which can potentially make significant contributions to testing procedures in university text-based studies of English. By focusing, in particular, on the concordancing of film texts and genres (Baldry, 2005), the article finds some answers to the question raised in the title. In so doing, it demonstrates how combinations of different types of concordance allow a few neat tricks, some old, some new, to be performed in the university teaching and testing cycle. The study begins by describing multimedia extensions to traditional *form-oriented language* concordances in relation to multimedia language tests (Section 2) and then illustrates the nature and use of genuine *meaning-oriented multimodal* concordances (Section 3). Finally, in keeping with its basic tenet of the need for multiple but integrated orientations in concordancing, the paper proposes a hybrid approach to testing in text-based university studies of English (Section 4) using a variety of concordance types.

2. FORM-ORIENTED CONCORDANCES IN RELATION TO LANGUAGE STUDIES BASED ON MULTIMODAL CORPORA

The screenshot shows the AntConc 3.2.1 interface. At the top, there is a menu bar with options: Concordance, Concordance Plot, File View, Clusters, Collocates, Word List, and Keyword List. Below the menu bar is a table of concordance hits. The table has two columns: 'Hit' and 'KWIC'. The 'Hit' column contains line numbers from 1 to 26. The 'KWIC' column contains the text from the script, with the word 'know' highlighted in bold in each line. Below the table, there are search controls. The 'Search Term' field contains 'know'. The 'Concordance Hits' field shows '26'. The 'Search Window Size' field shows '45'. There are buttons for 'Start', 'Stop', 'Sort', 'Save Window', and 'Exit'. At the bottom, there are options for 'Kwic Sort' with checkboxes for 'Level 1', 'Level 2', and 'Level 3', and input fields for '0', '24R', and '22L'.

Hit	KWIC
1	d in a sever beneath Vienna? as those of you know who saw the movie "The Third Man". Yes, that
2	ves. And I can recount all of them. How do I know? Very simple? because my name is Harry Lime.
3	h- GREGORY ARKADIAN: -Gregory Arkadian. You know who that is? HARRY: (chuckling in recogniti
4	so- GREGORY ARKADIAN: -Ha-ha! You think you know who is Gregory Arkadian! My friend, that is
5	rtugal- HARRY: -Really, old man, I wouldn't know- GREGORY ARKADIAN: -It's an allied operatio
6	Well that's nice offer, Arkadian. I have to know? who it is you want killed? before I give yo
7	are me a report. HARRY: What do you want to know about? GREGORY ARKADIAN: I want to know abo
8	to know about? GREGORY ARKADIAN: I want to know about? me. HARRY: You- GREGORY ARKADIAN:
9	ly? That's my real secret, Mr. Lime! I don't know who I am!!! MPX: HARRY: (narrating) So that
10	of a cinch. Oh there was plenty of cash, you know; unlimited traveling, expenses, and all of t
11	t to, but- hey-hey wait a minute, how do you know it's Arkadian- GREGORY ARKADIAN: -I know my
12	ou know it's Arkadian- GREGORY ARKADIAN: -I know my own name- HARRY: -Who says so- GREGORY A
13	ing a murder- HARRY: -Mr. Arkadian, I don't know how you began, but I know the best way to be
14	Arkadian, I don't know how you began, but I know the best way to begin finding out is- GREGOR
15	1? What did it say- HARRY: -Well he doesn't know; all he remembers is that the coat was made
16	ER CROWD BARONESS: I'm enjoying myself, you know. I didn't think I would- HARRY: -Go ahead,
17	e's been married since- HARRY: -Well do you know the man's name? BARONESS: I will look it up
18	gine you can trap me so easily- HARRY: -You know what, I bet you a couple of hundred bucks I
19	ysterious about everything Harry; you see, I know about the airports. There's to be three of t
20	t's why the intelligence check business- you know I don't even think he knows what an intellig
21	A: -He's heard people talking about it? you know Army people and others and- HARRY: -Yeah- R
22	ew up") -Hey? hey, just a second! How do you know he's Russian? RAINA: Well, I am his daughte
23	m his daughter- HARRY: -Yes, but he doesn't know himself where he was born! (RAINA begins to
24	: -Is that what he told you? That he didn't know where he was born? HARRY: You mean he was 1
25	Y ARKADIAN: -What about it- HARRY: -Well I know why you did it! You- GREGORY ARKADIAN: -Lim
26	crashing? I think because he wanted Raina to know? it wasn't accidental. Because he wanted her

Figure 1: Kwic sort of a radio script using AntConc 3.2.1

Within corpus studies, *form-oriented language concordancing*, in particular in the shape of KWIC [Keyword in context] concordances (Sinclair, 1991), has received most attention in classroom teaching. This type of concordance is instantly recognizable. The example in Figure 1, generated by Laurence Anthony's *AntConc* (Anthony, 2005), shows how the rows of individual concordances combine to produce a semi-tabular format with a single central column identified by automatically-created alignments, bold type, colour and gaps, allowing users to perceive patterns in wordings and to relate them to their co-texts (Sinclair, 1991). Concordances of this type greatly assist the phraseological approach (Hunston, 2006:55) to language studies. They are *monomodal* as they focus on retrieving entextualisations of *one* specific meaning-making resource from text corpora, namely language. They are also *form-oriented*. That is, rather than with the meanings made in texts, they are concerned with specifying how words form lexicogrammatical patterns, such as collocations and colligations (Hoey, 2006). A major reason for the widespread use of this type of concordancing is the availability of

lemma-based concordancers such as *Word Smith* and *AntConc* (see *Figure 1*) which identify typical patterns, some predictable, others less so, and the fact that tags, i.e. metatextual descriptions associated to specific forms, are either unnecessary or are efficiently embedded in texts thanks to automatic tagging systems such as *CLAWS* (Condron, et al., 2000).

In theory, non-linguistic form-oriented corpora and concordances can be produced relating to annotated and searchable sets of sounds with potential uses in language-learning and testing activities, e.g. *Write a short description outlining the actions, assumptions and expectations entailed by the sounds you will hear*. The sounds could relate to nature (*waterfalls*), animals (*wolves*), humans (*crying babies*), machines (*buzzers, phones, car engines*) and so on. In practice, with its numerous sub-branches and theoretical characterizations (Tognini-Bonelli, 2001), lemma-based concordancing seems to have eclipsed the possibility of research into other types of concordancing. Two questions thus need to be raised. First, whether approaches exist that can be realistically introduced into university text-based studies of English that are additional and/or complementary to traditional conceptions of lemma-based concordancing and, second, whether this will allow a greater focus on meaning-making processes in specific texts.

Classroom experimentation vis-à-vis films and related film scripts is a good starting point when attempting to provide an answer to these two questions. *Figure 2* illustrates three moments in a lesson recorded in the University of Pavia in January 2004, which prepared students for exams leading to B2-level international certification in English. Throughout this preparatory activity, the *media-indexed* concordancing facilities, which have existed in the online multimodal corpus authoring system, *MCA*, since its inception (Baldry, Beltrami, 2005), were used by the students to produce concordances which link viewings of scenes in a film to wordings in the relevant part of the film script, consisting of the characters' names and their lines. The students were asked to fill in a handout summarizing events in various phases in the film (for *phase* see Baldry, Thibault, 2006a: Chap. 4). They consulted each other, in English, as they reconstructed their understanding of the film on the basis of a single viewing of short fragments.

Part of this activity required the students to provide appropriate subtitles for a sequence of frames (*Figure 2.1*). To help them, the students were prompted by a series of phrases (*Figure 2.2*) and given access to computers to carry out keyword searches using *MCA* (*Figure 2.3*). This allowed them to establish how individual words, such as *work* and *job*, were used in the film script. It also gave them a chance to view and hear the specific film sequence – *utterance, subphase* or *phase* – in which these words occurred (*Figure 2.3*). Concordancing exercises were thus part of an overarching activity exercising various skills – reading, writing, listening, speaking and reflection on the functions of language in context – in an integrated way (Baldry et al, 2005).

In contrast to the traditional pattern-establishing activity of concordancing concerned with word distributions as described above, attempts were *also* made throughout this activity to base concordancing on the meaning-oriented principle of comparing *what might have been said* with *what was actually said* and *what might have been done* with *what was actually done*. Concordancing activity of this type stimulates students' understanding of what is going on in films, thus en-

1. Subtitling frames from the film

2. Phrases from the film script

Subphase 2c 136.5 190

Line 01 Line 1: Maggie: His name's Brian Whitman. He's been working here for eight years: video maintenance engineer. Suddenly blows a gasket... makes you wonder, huh?

Phase 7 439.5 512

Line 08 Line 8: Maggie: Well you've got to know! I mean, I've got a job, I just can't leave.

Reported (next) 2 rows, remaining 0

Search parameters

Line 01 [1468] Contains work

OR Line 08 [1475] Contains job

3. Concordancing with MCA: linking what was said with what was done

Figure 2: Stages in the reconstruction of meaning units in a film

couraging discussion about possible interpretations. What, for example, is the meaning of Maggie's exclamatory: *Well, you've got to know!* (Figure 2.3) in the context of the subsequent unfolding of the film text? The answer to this problem, and similar ones, can be obtained by an MCA search of the type *Lines contains "X"* where X is a character's name. More refined searches of the relational type: *Lines contains "Brian" AND Lines contains "Steven"* narrow down the search by identifying all the phases with lines uttered by both Brian and Steven, the latter being the person addressed by Maggie at this point in the film. This establishes that there are no *hits* i.e. Brian and Steven never meet up. A relational search of the type: *Lines contains "Brian" AND Lines contains "Maggie"* establishes, on the contrary, that Brian and Maggie *do* subsequently meet up, thus suggesting Maggie's duplicity when she says: *You've got to know* as opposed to *We've got to know*. It is now an easy step to work out that, unlike a fairy story, *what is said* and *what is done* in this film are out of step, leading students to form further hunches.

When applied to film or radio scripts, traditional word-distribution concordancing techniques throw up examples of the pattern *have to know/have got to know* (see Line 6 in Figure 1 and Line 8 in Figure 2.3) from which the meaning *must find out* can be deduced. However, except by roundabout methods, lemma-based concordancing in itself will not supply the answer as to *what will come to be known* and *how/whether/why/when/where it will come to be known* by the viewer in the context of the subsequent stages of this film. To achieve such a text-exploring function, concordancing must be adapted to the needs of university-level studies of texts by allowing and encouraging teachers and students to apply, critically and systematically, a meaning-oriented grammatical model, such as the one provided by systemic-functional linguistics (Halliday, 1994 [1985]). In the case in point, the concordancing procedure is embedded in classroom teaching and testing ac-

tivity which constitutes a first step in the exploration of the transitivity system in English; it establishes and discusses relationships between participants, their individual personality and their actions, real and potential, without needing to view the whole film first. Subsequent steps (see Section 4) are concerned with a fuller exploration of participant-process-circumstance relationships in terms of their direct embedding in concordances, in keeping with the goal of applying functional grammar through concordancing.

Despite some successes, there was a feeling throughout the classroom experimentation, shared by students and teachers alike, that textual explorations were being constrained by the limitations of traditional KWIC concordancing. Thus, lurking in the background was the constant, nagging question: how could patterns in wordings be better related to patterns of action? As Figure 3 shows, two further types of concordance were incorporated into MCA in July 2006: the *tabulated concordance* and the *captioned concordance* respectively associated with MCA's new *Pivot* and *Captioning* facilities. They constitute extensions to KWIC-type lemma-based concordancing and are designed primarily to take the co-presence of visual, verbal and actional resources in multimodal texts into account, encouraging users to speculate about the nature, incidence and functions of wordings in multimodal texts in terms of underlying meaning-making processes. Thus, the focus in the corpus illustrated in Figure 3 is on the way in which resources, visual, linguistic and actional, are used to build up clusters and their parts (for *cluster* see Baldry, Thibault, 2006a: 31). In this respect, this corpus, derived from a collection of printed posters produced for an Italian supermarket chain, instantiates two types of textual objects: *verbal clusters*, which function as captions, and *visual clusters*, which reincarnate foodstuffs as people, animals and inanimate objects and, in so doing, evoke potential actions. To reflect these realities, the corpus in question has been tagged as a *multimodal form-oriented corpus*. Thus, in keeping with traditional concordancing, searches can still focus on *wordings* but additionally there are tags for non-linguistic resources, both primary ones, such as colour, size, spatial disposition, and secondary ones such as the clusters that the primary resources instantiate: clothing, foodstuffs, humans, animals and objects. The combined force of the *media-indexed concordance*, the *tabulated concordance* and the *captioned concordance* means that a suitably tagged corpus can now be searched for non-linguistic data in ways that go beyond the *wordings-oriented* description given of these extensions in Figure 3.

RESEARCH	A captioned corpus			
Project Definition				
Play Media	start	end	Des1	Des2
Grammar Selection	0	4	Ballerina o carote?	A ballerina or carrots?
Media Indexing	5	9	Bosco o broccoli?	A wood or broccoli?
Task definition	10	14	Cavallucci o peperoncini?	Sea horses or chili peppers
Sequence Analysis	15	19	Colonne o formaggio?	Columns or cheese?
Search	20	24	Libellule o piselli?	Dragonflies or peas?
Pivot	25	29	Palle da tennis o pompelmi?	Tennis balls or grapefruit?
Captioning	30	34	Foche o melanzane?	Seals or aubergines?
	35	39	Mongolfieri o cipolle?	Hot air balloons or onions?



Figure 3: Examples of multimedia extensions to traditional lemma-based concordancing.

Extension 1: The media-indexed or multimedia concordance. By clicking on the *Media Player* buttons shown in the central panel, students can hear and/or see wordings as they occur in the original film sequence, in the case in point a sequence of digitalized posters. The corpus in question is multilingual and includes annotations in English, Italian and Albanian. As with traditional KWIC concordancing, the concordance shows how a target word – in this case Italian *o* – forms a recurrent pattern in relation to co-textual words that express ellipted *x-or-y* questions. In suitably tagged corpora, like the one illustrated, the *media-indexed* concordance, like traditional lemma-based concordances, can provide answers to such questions as to whether *x-or-y* questions in ellipted clauses in Italian and English (and maybe other languages) automatically involve a preferred choice that privileges the second of two options over the first. *Media-indexed* concordances thus help to create an awareness of the actual meanings that a specific pattern of *wordings* will make in specific contexts and specific textual genres in ways that apply but add to what can be learnt from book-based descriptions of grammatical systems. This is close to what is done in lemma-based concordancing but unlike traditional KWIC concordances, media-indexed concordances extend these possibilities to visually-oriented texts.

Extension 2: The tabulated concordance. The central panel shows this recent MCA extension which is designed to link *wordings* to the textual properties of specific genres and genre-lets. The tabulated concordance is implemented in MCA through the *Pivot* function which, as its name suggests, rotates traditional concordances through 90 degrees, presenting them as paradigmatic sets labelled with a classifying *Headword* (in this case the labels of convenience *Des1* and *Des2*). As befits an approach to concordancing focusing on text types, the row-based, semi-tabular format of traditional concordances, with its emphasis on a single *language form* and a *single column*; has given way to a fully tabular display consisting of columns containing sets of words with similar functions, in this case relating to the *caption* mini-genre. The example shows only two columns, but many more can be presented on a computer screen or printed out. As the central panel shows this extension to *wordings* encourages the examination of intertexts – in this case parallel subtexts in English and Italian consisting of captions and their translations. This furthers the possibilities for comparative text analysis. Like all expressions of intertextuality, *intertexts* are a way of looking at texts in intermediate terms, as entities lying half way between system and instance and between potential and actual (see Baldry, Thibault, 2006a:55). Here too, within a very traditional approach to concordancing, a small shift towards the examination of the relationship between meaning potential and action potential is being made.

Extension 3: The captioned concordance is yet a further extension designed to embed concordances into online teaching and testing applications. As exemplified in the right-hand panel, captioned concordances can be presented as subtitles generated in MCA by the combined use of the *Search*, *Pivot* and *Captioning* functions. Like traditional subtitles, they are synchronized with the unfolding visual text; the *wordings* displayed are, however, the result of user selections from any two columns in a *tabulated* concordance. They can thus be varied at will to show many *wording* combinations. Within the lemma-based approach to concordancing, one use is in the early stages of language learning where dual-language subtitling can facilitate the identification and meaning of words in an associated soundtrack.

As we move away from *monomodal* towards *multimodal form-oriented* concordancing, the question *What are concordances for?* begins to find answers in ways that are relevant to university text-based studies of English. At the very least, an increased range of text types can now be explored. However, a further and equally substantial consequence is the re-interpretation of the nature and function of *co-texts* in concordancing. A co-text (Sinclair, 1991) is no longer made up of just the word sets that immediately precede and follow target words as exemplified in *Figure 1*. As well as traditional co-texts, the concordancing mechanisms described above also generate *extended co-texts* (the film sequences, the posters etc.) as part of a concordance. In this respect, the co-text is now multimodal rather than monomodal. From another standpoint, however, the redevelopment of lemma-based concordancing may be taken as reconstruing concordances, and, in particular, their co-texts, as *subtexts of* and *points of access to* specific parts of digitalized texts (films, audio recordings, photos, printed media etc.). The description of the concordance extensions in *Figure 3* characterizes this process in some detail and explains, in particular, how the process of re-interpreting co-text is furthered by MCA's tabulating and captioning procedures. In keeping with the text-exploring function we have advocated above, concordances now function as *contextualizing subtexts* (Baldry, 2005: 97) of the texts to which they relate. That is, they directly assist the exploration of thematically-related sets of intertexts, e.g. *primary genres* such as captions and slogans; *phase types* such as typical scenes in films involving phoning, introducing oneself, ordering a meal or a drink and so on (Cocchetta, 2004). Through this contextualizing function, students are constantly invited to explore texts shifting their focus between actual realizations and the shared intertextual patterns that prototypically lie behind specific instances (Baldry, Thibault, 2006b: 177). Overall, within such intertextual patterns, the approach adopted allows and invites comparison, across languages and across modalities, of mini-genres and genrelets. This is particularly useful in text-based studies of English and motivates the development of tests that explore intertextual patterns (see *Section 4*).

To summarize, the new concordancing mechanisms introduced through MCA enact a shift from the traditional concern for word patterns in large corpora towards long-heralded computer-based explorations of specific texts and genres (Baldry, 1990; Baldry, Thibault, 2001). Once automatic tagging is put aside there is no reason why corpora and concordancers should not be constructed in a way that highlights recurrent *textual* patterns in specific texts. A defining feature of this approach to concordancing is the way information about a film can be quickly built up by traditional keyword searches coupled with viewings of short film sequences lasting on average between 15 and 30 seconds. The same principles and techniques can, of course, be applied to *corpora* of films and scripts that allow specific film genres and/or specific functions in films to be explored (Cocchetta, 2004). The research and development work into concordancing carried out in the *DIDACTAS* and *eColingua* projects in relation to multimedia language activities and tests (Cocchetta, 2004; Baldry et al, 2005, Dalziel, Metelli, *in press*) has thus led to new forms of lexis-oriented concordancing that have been applied *inter alia* to adverts, printed posters and as we shall see below (*Section 3*) to medical animations (Baldry, Guardamagna, *in press*). MCA's multimedia extensions could

also be used to explore other types of media, for example, the Harry Lime radio script and soundtrack shown in *Figure 1*.

3. MULTIMODAL MEANING-ORIENTED CONCORDANCES AND CONCORDANCING AS A TEXT MICROSCOPE

Though we have made a passing mention of *multimodal-form oriented concordancing*, all the illustrations given so far relate to lemma-based KWIC concordances – form-oriented monomodal concordances in our terminology – whose co-texts have been extended to embrace multimodal texts relating to printed and dynamic media. Thus, none of the concordances illustrated so far are meaning-oriented and most are only residually multimodal. At the opposite end of the spectrum, separated by an intermediate series of other types of concordance, such as *form-oriented multimodal* concordances and *meaning-oriented monomodal* concordances, lie *multimodal meaning-oriented* concordances whose conception is very different.

The bottom part of *Figure 4* gives an example of such a concordance produced by the fragment of an MCA mini-grammar shown in the top part of *Figure 4*. In keeping with the focus on meaning-making processes in film texts, the concordance explicitly identifies transitivity relations in both the video and soundtracks and, apart from the link to the relevant film clip, little else. Thus, this type of concordance makes no reference to form, whether linguistic or non-linguistic. It focuses exclusively on how soundtrack and videotrack interact by specifying how meaning processes are distributed across the two tracks; thus, the example specifies experiential meaning associated with the visual process of *binding* and links it to experiential meaning in a labelling function in the soundtrack. Thus, functional grammar is directly embedded in concordances (cf. O'Halloran, Judd's *Systemics* software, 2002).

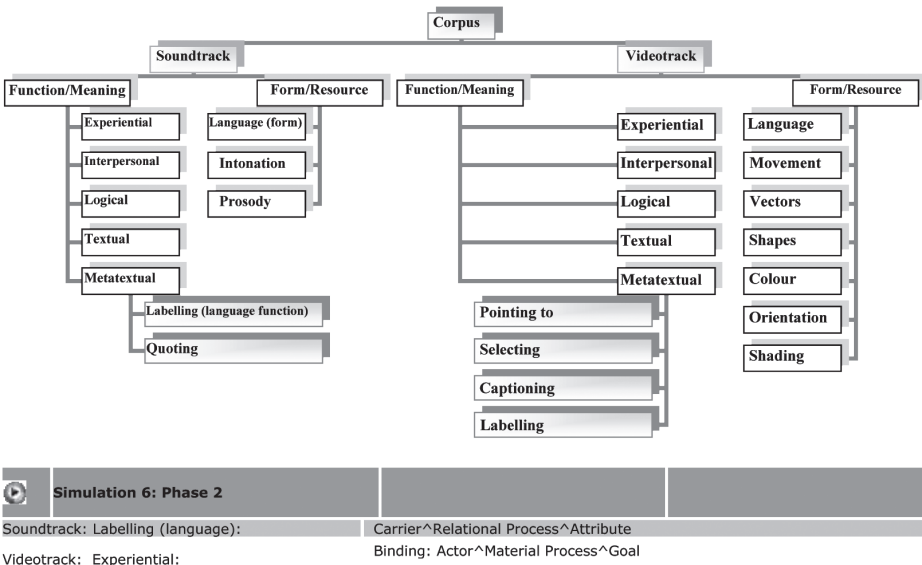


Figure 4: An MCA mini-grammar and a concordance that it produces

The concordance in question is, in fact, part of the exploration of the metatextual level of a corpus of 120 medical animations, designed originally for teaching purposes, that simulate and reconstruct body processes, so-called “mechanisms of action” such as heartbeat, which are affected by degenerative processes or by pathogens (Baldry, Guardamagna, *in press*). When exploring the medical animation genre, medical students, learning English through text-based studies, can use concordancing to establish how medical animations are organised as texts, in particular, the ways in which the off-screen narration is synchronized with what is happening in the visual. In general, the more complex the pathological processes become, the more complex is their multimodal representation and unfolding in time. This is the case, for example, with the videotrack’s metatextual level where visual processes such as captioning, labelling, selecting and/or pointing to organs and body parts are used to isolate and identify specific objects in the videotrack. A similar metatextual level exists in the soundtrack insofar as the off-screen narrator uses resources such as specific voice prosodies and salient intonation patterns in combination with language to carry out a labelling function that signals to the student viewer that a technical term is being used that needs to be remembered.

Interesting as the analysis of this corpus may be in its establishment of the nature and functions of *multimodal meaning-oriented* concordances, the fundamental point to be made in this section is that concordancing in MCA is adjustable to actual needs. It is the MCA user who decides the degree to which a particular set of concordances will be *monomodal* or *multimodal* and *meaning* or *form-oriented* by moving up, down and across the various branches of a mini-grammar tree. Let us illustrate this in relation to the mini-grammar given in *Figure 4* and the word “binding”. To obtain a traditional lemma-type concordance, the user selects the *Language (form)* descriptor (part of the *Form/Resource* subset of the *Soundtrack* branch of the mini-grammar) and types in the specific target word, in this case the word “binding”. A search of the form: *Language (form) contains “binding”* returns concordances containing all the examples of the word “binding” in the soundtrack together with their co-texts. The search can now be modified to include all the instances of the word “binding” that relate specifically to the labelling function. The search takes a relational form: *Language (form) contains “binding” AND Labelling (language function) contains “binding”*. This search returns a concordance which specifies both the wordings involved and their functions (in terms of transitivity relations). The user can now extend the search once again so as to produce concordances for all of the functions that the word “binding” carries out in the soundtrack. This time, the search form is: *Soundtrack (Function/Meaning) contains “binding”*. A further extension produces concordances that include *both* wordings *and* functions relating to “binding” in the entire soundtrack. In this case, the search takes one of two forms: either *Soundtrack (Function/Meaning) contains “binding” AND Soundtrack (Form/Resource) contains “binding”* or more simply, moving up the mini-grammar tree, *Soundtrack contains “binding”*.

However, the user can also go beyond the soundtrack to include annotations that describe the binding process in the videotrack. Binding is, after all, a basic visual process in this corpus with many fascinating “science-fiction” animations in which objects move and hook up with each other. Thus, although the word

“binding” does not appear in the videotrack, thanks to the way the corpus is annotated, a search of the form: *Videotrack contains “binding”* will produce concordances that identify all the visual processes involving binding. This includes data relating both to the *meaning-oriented* and *form-oriented* branches of the videotrack, i.e. both the process itself and the resources – typically movement, colour and shape – used to instantiate it. A more restricted search, one that is *either* meaning-oriented *or* form-oriented, will involve a subset that is lower down the tree, namely *Videotrack: Function/Meaning contains “binding”* or, alternatively, *Videotrack: Form/Resource contains “binding”*. There is a final expansion which produces concordances that include forms and functions of *all* references to binding – linguistic, visual and actional – in the entire corpus. The search is simply: *Corpus contains “binding”*.

In this way the user, starting from a low-level subset, has now reached the top of the mini-grammar shown in *Figure 4*. All this is a powerful illustration of the text microscope function of MCA concordancing. A concordancer now provides a level of focus and definition which can be constantly raised or lowered by the user. The reverse mechanism is also possible, i.e. gradual restriction from searches that involve both the videotrack and the soundtrack to more limited ones. Given that a film sequence is associated to all the concordances mentioned above and that the student is free to use media-indexed, tabulated or captioned concordances in all cases, the range of choice students have in their concordance-based text explorations is now considerable.

This degree of user control over concordance type, is possible because, as the chart in the top part of *Figure 4* shows, concordancing in MCA is governed by the specification of a *mini-grammar*, a level of annotation which defines descriptive parameters and which interacts with a second level of annotation – *Sequence Analysis* – in which specific data is associated to each descriptive parameter for each film sequence defined by the corpus author. In the case of the film described in *Section 1*, the mini-grammar is very simple, consisting of a single superordinate category, *Lines*, which contains a subset of descriptors of the form *LineX*, where *X* is a number corresponding to each of the lines uttered in a particular scene. In the case of the medical animation corpus, the mini-grammar in *Figure 4* is much more complex, consisting of a series of superordinate categories which contain subsets of descriptors, some of which, in their turn, function as superordinate categories containing further subsets. Thus, searching for *binding* as a visual process is technically possible because *binding* is represented in the MCA mini-grammar both in terms of the experiential functions instantiated in the videotrack and in terms of the resources used in their instantiation. As explained above, the details of the functions of the individual binding processes and the resources used to instantiate them are specified for each film sequence through the second level of annotation, namely the *Sequence Analysis* tool.

To summarize this section, we have described a concordancing procedure that extends traditional concordancing of language-based texts to multimodal texts in terms of a user-controlled sequence of graded stages. The focus has been on a variety of concordances including *genuine* multimodal concordances that identify basic text processes. Videotracks and soundtracks in the medical animations corpus and, of course, film texts in general, synchronize many meaning-making

processes. In this respect, we may look on sets of multimodal concordances as *records* of this activity arranged so that similarities and differences are highlighted *within* and *across* texts. That is, by rearranging activities so that their patterned nature is made more prominent, multimodal concordances reconstruct, in an ideal form, both *patterns* of relationships involving mergings between different activities and processes as well as *stages* in texts involving temporal and/or causal sequencing. Multimodal concordances thus identify similarities and differences in hierarchical terms, i.e. *within* the same type of activity and *within* the same text as well as in relational terms, i.e. *across* classes of activity and *across* texts. Obviously, but perhaps more significantly, they do so in terms of combinations of both. Though far from exhaustive, the mini-grammar in *Figure 4* illustrates the potential for different types of searches: monomodal/multimodal, meaning/form-oriented, relational and/or hierarchical. These options significantly empower the user as regards choice, turning a concordancer into a text microscope. All this constitutes a flexible approach to the analysis of meanings, forms and resources in a film corpus. *Section 4* shows why this degree of flexibility is essential when constructing concordance-based tests.

4. TESTS BASED ON A VARIETY OF CONCORDANCE TYPES

There is now only a short step to be taken as regards enacting a test that uses *inter alia* meaning-oriented concordancing in the assessment of students' awareness of the multimodal organisation of texts. The testing procedures illustrated in *Figures 2* and *3* are formative in that they relate to classroom activities. End-of-course, summative tests can also be devised that ascertain students' individual skills in understanding films based, for example, on the appropriate captioning of a sequence of frames as illustrated in *Figure 2.1*. *Figure 5* gives a checklist of questions used as the basis for such a test which may be considered "valid" for many multimodal genres. Specifically, by virtue of their reference to time-related events, the questions shown in *italics* could, in most circumstances, *only* be posed in relation to dynamic multimodal texts, such as film texts, whose meaning making is by definition dependent on the passage of time, while the majority of the other questions could also be put forward in relation to printed texts. In this respect, it may be claimed that, as hinted above, one of the main advantages of using multimodal concordances in relation to multimodal tests is precisely that they assist in the work of encouraging the use of film texts in language learning, something to be considered a big step forward.

- 1 Who/what are the participants in the text? Who/what are the main participants?
- 2 Are they human, animal or inanimate?
- 3 Which participants are active? Which participants are passive?
- 4 What relationships seem to exist between the participants? How do they interact?
- 5 What kinds of activities and events are associated to them? What temporal scales are involved?
- 6 What processes are undertaken and in what circumstances? What namings, highlightings, shadings and linkings exist? *What mergings, changes in size and perspective are found?*
- 7 *What features – e.g. patterns of movement, action, colour, sound – connect one part of the text to other parts? What changes take place in these patternings?*

8 What is the significance of colour? Do particular colours stand out? *What changes in colour as the text unfolds indicate special meanings?*

Figure 5: A checklist of questions applicable to multimodal texts

There are various ways in which the questions in *Figure 5* can be used to test students' awareness of the multimodal organisation of texts. We may briefly illustrate three.

A first type which has been successfully tested in the University of Pavia with students following *lauree specialistiche* in medicine and communication studies is for students to create a small MCA corpus of filmed websites usually restricted to two or three home pages and other pages directly linked to the home page that are tagged in such a way as to provide replies to these questions. Students are required in their final exam, to give a *PowerPoint* presentation that both illustrates and compares the characteristics of the chosen websites. At one level, this is a test of language ability that specifically requires students to integrate written content with an oral presentation. At another level, it is a test of students' awareness of multimodal grammar (Baldry, Thibault, 2006a). Concordances of various types, derived from the students' mini-corpora, provide the basis for tabulated comparisons in the *PowerPoint* presentation of the various mini-genres in the websites selected for comparison, such as search engines, mastheads, photos, captions and so on. Tabulated, multimodal form-oriented concordances (described in *Section 2* in relation to the poster corpus) relating to cluster analysis are popular in the preparation for this kind of presentation. This is partly because they help students to identify the shared, intertextual characteristics and functions that lie behind the many variables, functional and formal, associated with these mini-genres.

A second test involves the medical simulation corpus described in the previous section and is based on the instruction to students enrolled in the Faculty of Medicine to write a summary in English relating to the characteristics of the medical simulations genre, including typical transitivity patterns. This type of test requires students to carry out multimodal meaning-oriented searches in relation to the typical phenomena of this genre: binding, breaking, transforming and so on.

A third test involves the printed poster corpus. This requires students to write a guided summary of the poster genre based on the corpus shown in *Figure 3*. Though similar to the preceding test, the corpus is tagged in such a way as to provide the students with an important shortcut. As explained in *Section 2*, the corpus is based on a mini-grammar, part of which incorporates descriptors of the type indicated in *Figure 4*, but part of which also includes sets of questions like those indicated in *Figure 5*. Searches thus can be carried out directly in terms of test questions such as those in *Figure 5*, all part of the process of making the text-exploration function of corpora easy for students to carry out.

To summarise, this section has focused on tests based on a variety of concordance types for the analysis of multimodal genres that this article proposes as being appropriate in a teaching and testing cycle. This includes genuine *multimodal* tests concerned with *explicit* demonstrations of students' awareness of grammatical functions in multimodal texts (Baldry, Thibault, 2006a).

5. CONCLUSIONS

What are concordances for? One answer to this question, the main one being put forward in this article, is that concordances need not necessarily be concerned *merely* with the quest for patterns in texts. A different function of concordancers is to pinpoint words and other non-linguistic resources in text-based corpora and to characterise their contextual functions. A further concomitant function is to allow texts to be explored, reconstructed and analysed. We may recall that with the development of computer-based concordancing, a shift occurred as regards the nature and function of concordances as compared with their original incarnation as printed word lists which helped scholars locate words and phrases in literary texts (e.g. concordances to Shakespeare). The MCA project returns to this original conception by pinpointing the occurrence of specific words and other resources in specific texts. But it goes beyond this by making it possible to describe the functions of these words and resources in terms, for example, of the transitivity system in English, by allowing characterisations to be made of the participants, human or otherwise, the processes they enact and undergo and the circumstances in which these processes occur. Classroom experimentation has shown this approach to be a worthwhile venture.

Is it difficult to implement? Since its inception, MCA has been designed to be easy to use. The concordancing activities described above only require the selection of a descriptor from a drop-down menu, a keyword to be typed in and the *Search* key to be activated. From an author's point of view, the task is not quite so easy since manual tagging is involved. However, description of the different concordance types and concordancing procedures that this article has provided is an essential step to making the authoring goals clearer to understand and possibly simpler to implement. In keeping with this principle, the article has redefined co-text and suggested that four different types of concordance exist, namely:

- (1) Monomodal form-oriented concordances (the traditional lemma-based concordance);
- (2) Monomodal meaning-oriented concordances;
- (3) Multimodal form-oriented concordances;
- (4) Multimodal meaning-oriented concordances.

Furthermore, it has associated each of these with three new concordancing procedures namely:

- (1) The media-indexed concordance;
- (2) The tabulated concordance;
- (3) The captioned concordance.

This does not, of course, exhaust the list of possible concordance types or the concordancing procedures associated with them. It is possible, for example, to posit the existence of the scalar-oriented and genre-oriented concordances as significant subcategories of multimodal meaning-making concordances (Baldry, Thibault, forthcoming).

As MCA evolves, it attempts to transcend the possibilities afforded by *form-oriented language-only* concordance tools essentially concerned with the concordancing of words, their parts and combinations such as: SARA (Aston, Burnard, 1998),

TACTweb (Rockwell et al., 1997), *WordSmith Tools* (Scott, 2001) and *AntConc* (Anthony, 2005) while still retaining and incorporating, where possible, the principles and affordances inspiring them. Once automatic tagging is put aside there is no reason why corpora and concordancers cannot be constructed in such a way as to support all the types of concordancing discussed above. As more experience with the media-indexed, tabulated and captioned varieties of monomodal and multimodal form-oriented concordancing is acquired, the time taken to complete the tagging process for these types of concordancing will constantly be reduced. Faster tagging is expected with the next release of MCA which will include file-uploading tools that extend the possibilities for offline tagging of files based on word processing tools such as *Word*.

The search for a *genuine* multimodal approach to concordancing – one that is not simply lemma-based concordancing re-invented with multimedia extensions that produce a multimodal co-text (*Section 1*) – has been particularly influenced by observations expressed by researchers, teachers and, of course, students about the goals of computer-based concordancing. Thus, the research reported above is breaking new ground not just by promoting new genuinely multimodal forms of concordancing and new concordancing techniques. Rather, within *applicative* frameworks, it is attempting to define new multimodal goals for concordancing that use concordances to apply functional grammar to multimodal texts. The article has suggested the viability of alternatives, developed and tested within the *DIDACTAS* and *eColingua* projects, to traditional lemma-based concordancing, with students taking courses in the fields of foreign languages, communication studies and medicine. All this is part of a constantly evolving process whose developments include: research into new forms of lexis-oriented concordancing appropriate for the classroom analysis of visual genres (Baldry et al, 2005; Ackerley, Coccetta, *this volume, in press*; Dalziel, Metelli, *in press*); probabilistic approaches to form-oriented language-only concordances (Tucker, 2006: 90-4); the development of language-based meaning-oriented tagging (Taylor Torsello, 2001, Taylor Torsello, Baldry, 2005); language-based meaning-oriented concordancing as implemented by *Systemics* (O'Halloran, Judd, 2002); and research into multimodal corpus linguistics (Baldry, Thibault, 2001, 2006b, *forthcoming*).

To summarize and conclude: a basic tenet that has informed this paper is that a wider vision of the nature and functions of concordancing than has existed in the past is, *inter alia*, a key to better integration of concordances into the university teaching and testing cycle. Testing, in particular, is an area where corpus studies appear, so far, to have been used only in a limited and experimental way (Barker, 2004). As we have seen in *Section 4*, new forms of concordancing that use the MCA system can contribute to the goal of underpinning the position of multimodal text analysis in the university teaching and testing cycle by strengthening the range and variety of classroom testing activities based on concordancing. In a nutshell, by advocating a step-by-step approach that blends traditional concordancing with more innovative conceptions, the article has posited a multifaceted answer to the question raised in the title.

SOFTWARE

AntConc: www.antlab.sci.waseda.ac.jp

CLAWS, the Constituent Likelihood Automatic Word-tagging System, UCREL, University of Lancaster, 1987. <http://www.comp.lancs.ac.uk/computing/research/ucrel/history.html>.

MCA: Multimodal Corpus Authoring System: <http://mca.unipv.it>.

SARA, SGML-Aware Retrieval Application, Humanities Computing Unit, Oxford University Computing Services, 1994; <http://www.hcu.ox.ac.uk/BNC/using/papers/burnard96a.htm>

Systemics 1.0., O'Halloran, Kay L. and Judd, K., 2002. [CD-ROM]. Singapore: Singapore University Press. <http://courses.nus.edu.sg/course/el-lkoh/Overview.html>

TACTweb Bradley, John and Geoffrey Rockwell, 1997. <http://tactweb.humanities.mcmaster.ca/>

WordSmith Tools, Scott, Mike, 1996. Oxford: Oxford University Press, version 1.0, 1996 to v. 3.0, 1999, & v. 4.0 forthcoming. <http://www.lexical-ly.net/wordsmith/>

BOOKS AND ARTICLES

Ackerley, Katherine, Cocchetta, Francesca (this volume). "Multimodal concordancing for online language learning: exploring language functions in authentic texts".

Ackerley, Katherine, Cocchetta, Francesca (in press), "Multimodality in an online English course", in: Anthony Baldry and Elena Montagna, (eds.), *Interdisciplinary Perspectives on Multimodality: Theory and practice. Proceedings of the Third International Conference on Multimodality*. Campobasso: Palladino.

Anthony, Laurence (2005), "AntConc: design and development of a freeware corpus analysis toolkit for the technical writing classroom". *Professional Communication Conference, 2005. IPCC. Proceedings. International*, 729-37.

Aston, Guy, Burnard, Lou (1998), *The BNC handbook: exploring the Bri-*

tish National Corpus with SARA. Edinburgh, Edinburgh University Press.

Baldry, Anthony (1990), *Research into Self-Access Study for Advanced Learners: Computer-Assisted Studies of English*. Udine, Campanotto Editore.

Baldry, Anthony (2005), *A Multimodal Approach to Text Studies in English. The role of MCA in multimodal concordancing and multimodal corpus linguistics*. Campobasso, Palladino.

Baldry, Anthony, Beltrami, Michele (2005), "The MCA Project: concepts and tools in multimodal corpus linguistics", in: Maj Asplund Carlsson, Anne Løvland and Gun Malmgren (eds.), *Multimodality: Text, culture and use. Proceedings of the Second International Conference on Multimodality*. Kristiansand, Agder University College/Norwegian Academic Press, 2005, 79-108.

Baldry, Anthony, Caldirola, Elena, Grunther, Patti (2005), "Knowing and learning about what to expect: The role of multimodal concordancing in explorations of film texts in foreign language short courses", in: Christopher Taylor, Carol Taylor Torsello e Maurizio Gotti (a cura di), *Centri Linguistici: approcci, progetti e strumenti per l'apprendimento e la valutazione. Atti del 3° convegno nazionale AICLU*. Trieste, Università degli Studi di Trieste, Dipartimento di Scienze del Linguaggio, dell'Interpretazione e della Traduzione, 547-63.

Baldry, Anthony, Guardamagna, Caterina (in press), "Medical animations: their place in multimodal corpora and concordancing", in: Anthony Baldry and Elena Montagna (eds.), *Interdisciplinary Perspectives on Multimodality: Theory and practice. Proceedings of the Third International Conference on Multimodality*. Campobasso, Palladino.

Baldry, Anthony, Montagna, Elena (eds.) (in press), *Interdisciplinary Perspectives on Multimodality: Theory and practice. Proceedings of the Third*

- International Conference on Multimodality*. Campobasso, Palladino.
- Baldry, Anthony, Thibault, Paul. J. (2001), "Towards multimodal corpora", in: Guy Aston and Lou Burnard (eds.), *Corpora in the Description and Teaching of English*. Bologna, CLUEB, 87-102.
- Baldry, Anthony, Thibault, Paul. J. (2006a), *Multimodal Transcription and Text Analysis. A multimedia toolkit and coursebook*. London and New York, Equinox.
- Baldry, Anthony, Thibault, Paul. J. (2006b), "Multimodal corpus linguistics", in: Geoff Thompson and Susan Hunston (eds.), *System and Corpus: Exploring Connections*. London and New York, Equinox, 164-83.
- Baldry, Anthony, Thibault, Paul. J. (forthcoming), *Multimodal corpus linguistics*. London, Routledge.
- Barker, Fiona. (2004), *Using corpora in Language Testing*, in: "Modern English Teacher", 13/2, 63-7.
- Cocchetta, Francesca (2004), *Language Functions in Films: A multimodal corpus for the English language teaching*. Padua: Unpublished degree thesis, Facoltà di Lettere e Filosofia, Dipartimento di Lingue e Letterature Anglo-germaniche e Slave, University of Padua.
- Condron, Frances, Fraser, Michael, Sutherland, Stuart (2000), *CTI Textual Studies guide to digital resource for the humanities*. Oxford, Humanities Computing Unit University of Oxford.
- Dalziel, Fiona, Metelli, Marianna (in press), "Learning about genre with a multimodal concordancer", in: Anthony Baldry and Elena Montagna (eds.), *Interdisciplinary Perspectives on Multimodality: Theory and practice. Proceedings of the Third International Conference on Multimodality*. Campobasso, Palladino.
- Halliday, Michael (1994 [1985]), *An Introduction to Functional Grammar. 2nd edition*. London and Melbourne, Arnold.
- Hoey, Michael (2006), "Language as choice: what is chosen?", in: Geoff Thompson and Susan Hunston (eds.), *System and Corpus: Exploring Connections*. London and New York, Equinox, 37-54.
- Hunston, Susan (2006), "Phraseology and system: a contribution to the debate", in: Geoff Thompson and Susan Hunston (eds.), *System and Corpus: Exploring Connections*. London and New York, Equinox, 55-80.
- O'Halloran, Kay (ed.) (2004), *Multimodal Discourse Analysis*, London and New York, Continuum.
- Rockwell, Geoffrey, Passmore, Graham, Bradley, John (1997), *TACTweb: The Intersection of Text-Analysis and Hypertext*, in: "Educational Computing Research", vol. 17, no. 3, 217-30.
- Scott, Mike (2001), "Comparing corpora and identifying key words, collocations, and frequency distributions through the WordSmith Tools suite of computer programs", in: Mohsen Ghadessy, Alex Henry, Robert. L. Roseberry (eds.), *Small corpus studies and ELT: theory and practice*. Amsterdam, John Benjamins, 47-67.
- Sinclair, John (1991), *Corpus, Concordance, Collocation*. Oxford and Singapore, Oxford University Press.
- Taylor, Christopher J., (2004), "Multimodal text analysis and subtitling", in: Eija Ventola, Cassily Charles and Martin Kaltenbacher (eds.), *Perspectives on Multimodality*. Amsterdam, Benjamins, 153-72.
- Taylor Torsello, Carol (2001), "Tagging systemically". Paper presented at ISFC XXVIII, Carleton University, Ottawa.
- Taylor Torsello, Carol, Baldry, Anthony (2005), "SFL in text-based, web-enhanced language study", in: Ruqaiya Hasan, Christian Matthiesen and Jonathan Webster (eds.), *Continuing Discourse on Language. A functional perspective. Vol. 1*. London and New York, Equinox, 311-42.
- Tognini-Bonelli, Elena (2001), *Corpus Linguistics at Work*. Amsterdam, Benjamins.
- Tucker, Gordon (2006), "Systemic incorporation: on the relationship between corpus and systemic functional grammar", in: Geoff Thompson and Susan Hunston (eds.), *System and Corpus: Exploring Connections*. London and New York, Equinox, 81-102.