Applied Linguistics 2016: 37/1: 7–32

doi:10.1093/applin/amv065

Applying Digital Sensor Technology: A Problem-Solving Approach

[1,*]PAUL SEEDHOUSE and [2]DAWN KNIGHT

[1]School of Education, Communication and Language Sciences, Newcastle University and [2]School of English, Communication and Philosophy (ENCAP), University of Cardiff
*E-mail: paul.seedhouse@ncl.ac.uk

There is currently an explosion in the number and range of new devices coming onto the technology market that use digital sensor technology to track aspects of human behaviour. In this article, we present and exemplify a three-stage model for the application of digital sensor technology in applied linguistics that we have developed, namely, Technology–Problem–Iterative Development and Research. We present three projects that have used this model. In the first and second, a language learning environment was facilitated and tracked by digital sensor technology, while in the second and third projects, the technology enabled multimodal data collection and analysis. All projects investigated how a digital learning environment might be designed, implemented, and evaluated. The research focus has been on how to record and analyse the process of language learning through spoken interaction using digital sensor technology. This model is amenable to a variety of methodological approaches, as we see conversation analysis used in the first two projects and multimodal corpus linguistics in the third.

INTRODUCTION

The research presented in this article was developed in iLab:Learn, a laboratory for developing appropriate educational applications of digital technology at Newcastle University http://di.ncl.ac.uk/ilablearn/. The installations housed in iLab:Learn include a variety of multi-touch and pen-based tabletops, wearable tracking devices, and an instrumented kitchen for task-based language learning. The idea of an 'iLab' came about through the Digital Institute's initiative to support and promote the research, development, and application of digital technologies within Newcastle University's research. iLab:Learn grew out of collaborations between academic staff in the school of Education, Communication and Language Sciences and those in the school of Computing Science and aims to develop a distinctive program of technology-enhanced learning research that exploits social computing, pervasive computing, and situated interaction technologies and applications. This program has entailed developing a distinctive approach to research that is able to complement the technological innovations. This article presents the approach to research that we have developed, with an emphasis on the research methods involved. The relationship between theory and practice in research, technology, pedagogy, and interaction is a very complex one. We attempt to portray this relationship through narratives of how projects have unfolded and by using our three-stage model for application of digital sensor technology, namely, Technology–Problem–Iterative Development and Research.

First, we try to understand the potential of the digital technology, namely, what it can and cannot do. Secondly, we identify an existing and worthwhile problem in the field of applied linguistics that may be tackled using the technology. Thirdly, we engage in iterative development and evaluation of the pedagogical and technological system design with human subjects. We revise the design as we receive evidence of system use. We have found that a problem–solution approach is most appropriate to this specific area of research. Research design has to focus not only on product or outcomes in terms of language learning, but also on the process of user engagement with the digital learning environment. This is accomplished by a combination of automatic logging of digital sensor activity (who does what, when) and video and audio recording of user spoken interaction, which is transcribed for conversation analysis (CA), thus providing access to the micro detail of behaviour. There are two purposes in our digital research; first, to provide data for iterative system re-design, and secondly, to tackle the problem identified at the start.

We briefly outline three projects that were undertaken using different types of digital sensor technology. The first two projects involved foreign language teaching (French and English) and, specifically, implementation of approaches to task-based language teaching (TBLT). In the first project, we tackled the problem of motivation for L2 learning in the UK by building a Digital Kitchen in which users can learn French language, cuisine, and digital skills at the same time. In the second project, we used digital tabletops to tackle the problem of how task-based interaction for language learning can be analysed. In the third project, we questioned the potential for repurposing the sensors used in the Digital Kitchen study as a means of tackling a different problem: the potential for using these devices as a means of capturing, encoding, and analysing patterns of language and gesture-in-use for the integration in multimodal corpora.

THE FRENCH DIGITAL KITCHEN PROJECT

In this section we outline, using the Technology–Problem–Iterative Development and Research model outlined above, the nature of the technology encountered, how we applied it to a problem in applied linguistics, how the learning environment was designed and researched, and we look at an example of it in action.

Technology

First, we familiarised ourselves with an existing technologically enhanced kitchen known as the 'Ambient Kitchen' (http://www.ncl.ac.uk/ihs/research/project/2756). This was originally developed by human–computer interaction scientists at Culture Lab at Newcastle University to support older people and those with dementia in their everyday kitchen activities. The term 'ambient' refers to the nature of the technology used in the kitchen that is absorbed or hidden in that environment and, similar to a car satellite navigation system, is designed to guide and support the user in an everyday setting. As an assistive-technology setting, the 'Ambient Kitchen' was designed to provide situated support in the form of written or audio prompting during a kitchen-based activity such as cooking or making a cup of tea. It did this by using digital sensors to detect actions and linking these to the possible intentions of the user; for example, filling the kettle could be a prelude to making a cup of tea. When we familiarised ourselves with this technology, it became clear that this could be adapted to the field of language learning and would fit easily into a task-based learning approach, in terms of learning through cooking.

Problem

But what would be the rationale and which problems might be addressed by such an approach? In the UK, recent years have seen a significant decrease in the number of schoolchildren choosing to study foreign languages at secondary school, which has implications for the broader economy, especially for commerce, tourism, and research. A number of research projects have therefore tried to develop new ways of engaging the UK population (at all ages) with language learning. So, adapting the Digital Kitchen to language learning while cooking a foreign dish offers the opportunity of taking the excellent research-based pedagogical principles and procedures developed by TBLT over the years out of the classroom and into use in real-world applications. There are a number of well-known problems relating to classroom foreign language teaching in the UK that were addressed by this project. First, the universal problem of classroom language teaching is that students are rehearsing using the language, rather than actually using the language to carry out real-world tasks. This problem has already been addressed by TBLT, but this project takes TBLT principles into a kitchen environment and has the learners learning a foreign language while actually cooking a foreign dish. Secondly, there is the difficulty of bringing the foreign culture to life in the classroom. In the Digital Kitchen, learners are able to learn aspects of the language (e.g. vocabulary items) while performing a meaningful real-world task and simultaneously experience the cultural aspect of learning to cook a foreign dish. The third problem is the lack of motivation for learning foreign languages felt by many British people. The project was therefore intended to tap into two other strong current motivations or interests of the British population, namely, cooking and technology. As

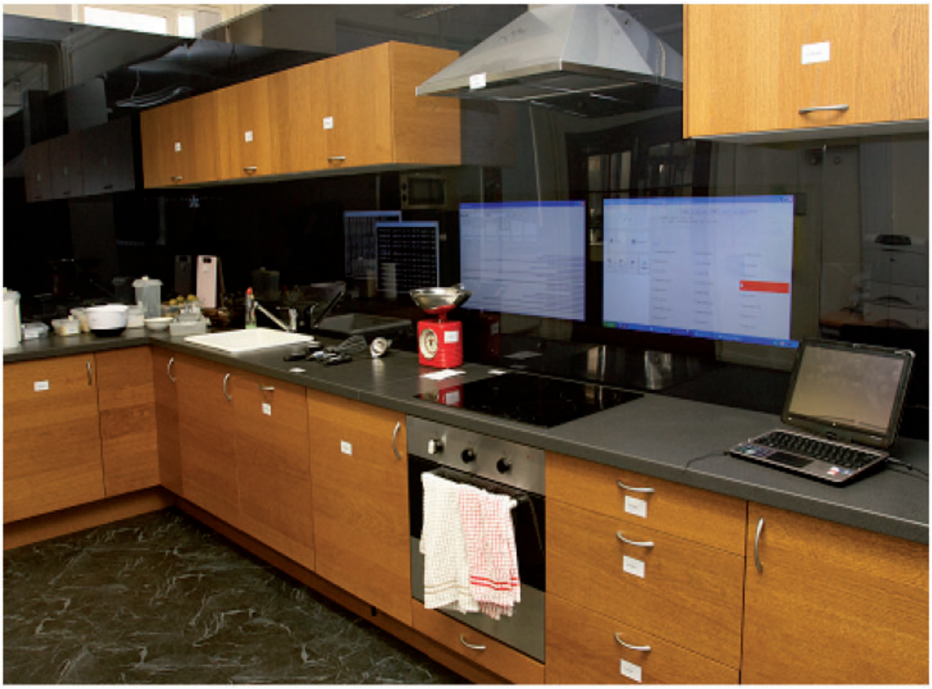


Figure 1: Purpose-built French Digital Kitchen

French has traditionally been the international language of cuisine, and has been widely taught in British schools, it was the most suitable language for the Digital Kitchen.

This project therefore aimed to create a situated language learning environment in which the kitchen speaks to the users in French, instructing them step-by-step in how to cook cuisine and helping them learn aspects of the French language. The kitchen attempts to combine learning, technology, and cuisine to develop motivation and interest in French language, cuisine, and culture.

This project took the technology of the existing Ambient Kitchen for communication with people with dementia and adapted it to the field of learning French language and cuisine. We constructed a purpose-built French Digital Kitchen (Figure 1) that speaks to the learners in French, providing step-by-step cooking instructions in relation to learners' completion of the cooking steps. It can also detect what the learners are (or are not) doing and this information is used by the kitchen program to provide feedback such as a reminder or more details about a certain cooking action in French. Using technology and a 'real world' task-based approach, the French Digital Kitchen offers a further way of promoting communication in French and engagement with French culture and cooking.

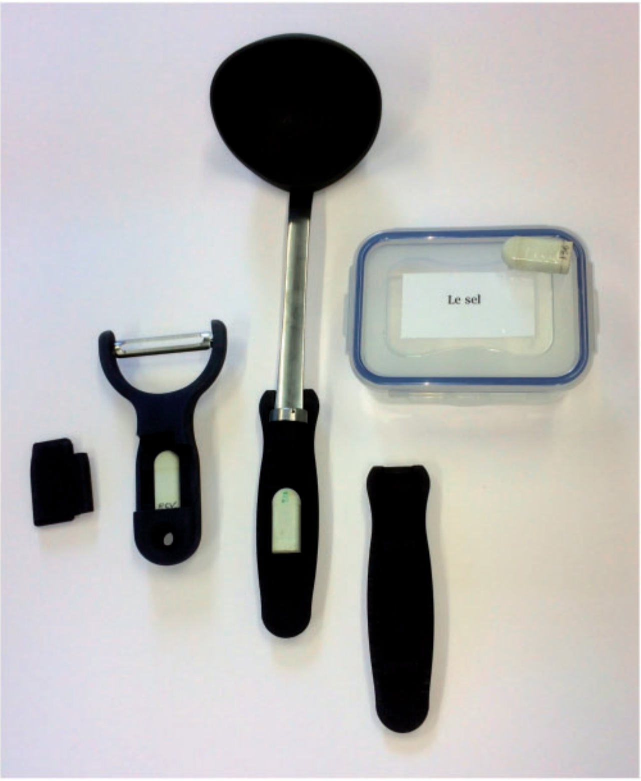


Figure 2: Sensors attached to ingredients and utensils

Iterative design: technology

Accelerometer sensors that detect three-dimensional movement are attached to all ingredients and kitchenware (Figure 2) so that each time an item is correctly or incorrectly moved, verbal feedback is given to the participants. The sensors use a technology similar to the Nintendo Wii™. The sensors hidden in the knife, for example, were designed to detect whether a 'chopping' action' or a 'scraping' motion is being made and provide appropriate feedback. The program moves through the cooking instructions step-by-step as it receives evidence from the sensors regarding the actions that the participants have carried out in relation to the stages of the task. In a similar way to a car's satellite navigation, the system provides feedback to users on their actions, for example, by explicitly informing them that they have performed an action correctly or incorrectly. If users do not understand the instructions in French, they are able to request repetition or translation of the instructions using an interactive touch screen (Figure 3). The project provides an example

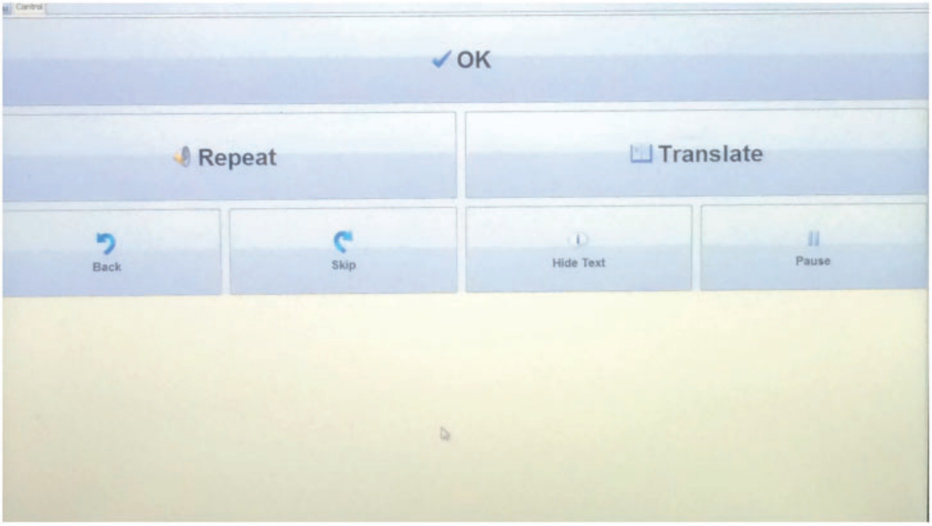


Figure 3: Interactive screen

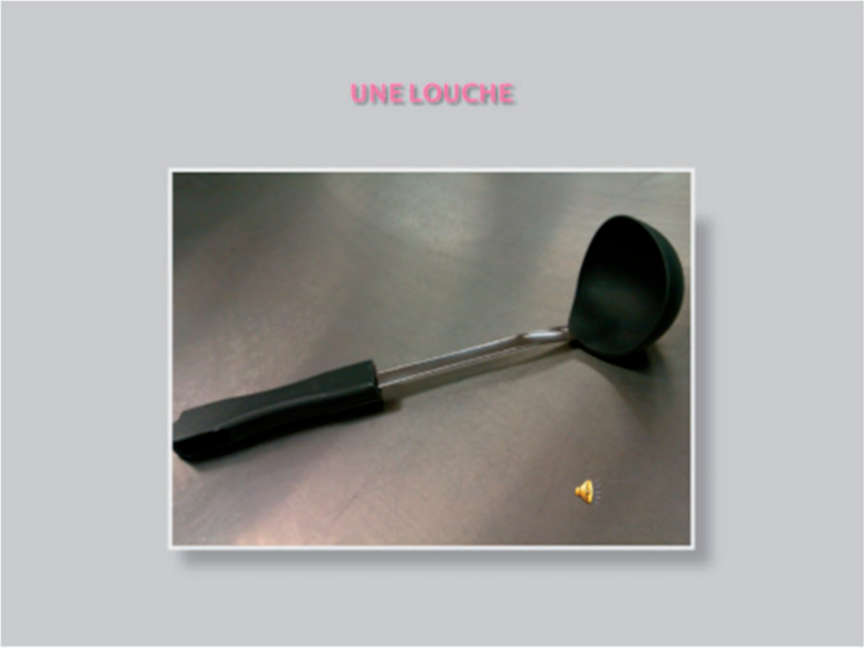
of how two rather different sets of skills may be acquired at the same time by use of appropriate technology. In addition, users are also developing digital literacy skills by learning to interact with the system.

Iterative design: pedagogy

Above we have provided a brief description of how the technology works. We now explain how pedagogical design and technological design were integrated to create a functioning digital environment for language learning. The project also aimed to promote learning of culinary skills and of digital literacy as well, but in this section, we focus on design for language learning. We wanted to develop an approach that could encompass the full range of pedagogical principles and procedures developed by TBLT (Skehan 1998; Ellis 2003) over the years but to implement these outside the classroom in a real-world setting, namely, a kitchen. Given the emphasis of the authentic task within TBLT, we have used the kitchen environment as a learning context, as the act of cooking a meal is an authentic task with a clear goal and tangible end product. Additionally, some people learning foreign languages are often motivated by a desire to immerse themselves in a foreign culture and cuisine: cooking feeds into this motivation. Instructions, reminders, and other language support are given in French and, when necessary, in English. That way we transform a regular kitchen environment into a TBLT tool where learning the foreign language happens naturally as the students cook. The relationship between TBLT and second-language acquisition processes is detailed by Long (2015). In TBLT, a task is designed to ensure that meaning is primary, there is a communication problem to solve, there is a relationship to comparable real-world activities, completion has some priority, and assessment is in terms of outcome (Ellis 2003). The underlying task in the kitchen was designed to encourage learners to focus on meaning rather than language alone. Nonetheless, incidental focus on form is available, using the supports detailed below. Secondly, learners must draw on their language skills to achieve the task. Thirdly, the task is situated in an authentic real-world context. In addition, the task is goal-oriented: it is clearly defined and has a goal. Pairing participants promotes communication in L1 or L2 while the task is being completed, as illustrated in Extract 1 below. We paired learners with French skills with those with culinary skills to create an information gap, so that they might transfer skills to each other. In classroom TBLT, the teacher is available for help, while in the Digital Kitchen, help is available via the interface (Figure 3) in terms of repetition or translation of instructions. The technology described above offers the advantage over some classroom-based TBLT that learners receive immediate feedback on whether they have carried out task-relevant actions correctly or not.

The cooking session was designed (adapting Skehan 1998) with three main phases: pre-task, cooking task (main task), and post-task. The pre-task combined a focus on cooking and French skills and involved presentation and preparation stages for both French and cooking. First, learners watched a purpose-made video recording of a native-French speaker making the chosen dish for the project, *Clafoutis aux poires*. Learners had a choice of watching this without subtitles, with French subtitles, or with English subtitles, depending on their levels of French and culinary proficiency and the level of support they required. Next, the learners watched an audio-visual slideshow of the different utensils and ingredients needed to make the dish, to familiarise them with the vocabulary required, as in Figure 4.

These activities were displayed on specially designed digital display screens on the walls of the kitchen (Figure 1). The final stage of the pre-task involved the Kitchen saying what ingredients were required and how much. The learners had to note these down and had the opportunity to request help, such as a translation in English or the repetition of a phrase. The cooking task itself involved the users following instructions of how to prepare the dish, aided by a range of relevant feedback provided by the Kitchen as and when needed, prompted by the learners' actions. Feedback included creating alternative versions of instructions, often reformulated in terms of 'tips' about cooking technique, which acted as prompts. English translations were also created using cooking-specific vocabulary. In the post-task, targeted vocabulary was evaluated using recognition tests on the display screens in the kitchen. The dish produced by the learners (the task outcome) was evaluated by being eaten by the learners and the researchers, who also interviewed the learners about their experiences and their learning. Whereas classroom-based TBLT may engage the learners' senses in terms of sight, sound, and touch, the Digital Kitchen engages the senses of smell and taste as well, delivering a vivid, kinesic language learning



UNE LOUCHE

Figure 4: Example of slide in audio-visual slideshow

experience (Seedhouse 2015). There is added value in performing the meaningful, embodied task of food preparation, which is common to all human cultures.

Illustration

To illustrate what actually happens when users carry out the cooking task in the French Digital Kitchen, we will examine some interactional data. We paired students with higher level of catering skills with students with higher levels of French in the expectation that they might be able to transfer skills to some extent. S1, a catering expert with little knowledge of French, with S2, an upper intermediate learner of French with little knowledge of catering. Translations are italicised.

Extract 1

1 KIT: mélangez ces ingredients (*mix these ingredients*)
2 S2: mix them together
3 KIT: et réalisez un puits dans la farine (*and make a well in the flour*)
4 S2: when you mix them (.) you've got to make a hole in the centre=

Extract 1

5	S1:	=a well in [the cen]tre
6	S2:	[yeah]
7	S1:	yeah (2.1) need to break them up
8		(2.1)
9	S2:	is that
10	S1:	yeah they'll break up (.) if you just give em a (1.0) .hh mix around
11		(1.2)
12	S1:	what's mix (.) in F-
13	S2:	mélanger (*mix*)
14	S1:	mélanger (.) mélanger oeufs (*mix eggs*)
15		(1.5)
16	S2:	and then to make (0.4) faire un trou au milieu (.) (*make a hole in the middle*)
17		is another way of saying what do you do to them
18	S1:	u::m
19	S2:	un trou (.) a [hole]
20	S1:	[a] well
21	S2:	a kind of (.) hole

In this extract we see how the task involves a dual orientation to the cooking task and to language. In this case, there is a tension between the two orientations. In lines 1 and 3, we hear the system giving the students instructions in French on how to proceed with the cooking, which S2 translates into English for S1's benefit. It is important to understand that the technical catering term to use in the case of making a hole in a quantity of flour in which eggs will be poured is 'puits' in French and 'well' in English, which is a literal translation. So although S1 may not understand the instruction in French in line 3, it appears he knows from the context of the cooking operation that they are making a well and uses that technical term in line 5, and again in line 20. S2 does not appear to know the technical catering term in French or English and refers to it as 'hole' in lines 4, 19, and 21 and as 'trou' in lines 16 and 19. S2 appears to be trying to teach S1 the French word 'trou' in 16 and 19, although this was not a word spoken by the Digital Kitchen. However, the point is that the participants display such an orientation to completing the cooking task that some confusion over linguistic terms is not a problem, given the context of the hole in the flour that they have created. In lines 7–11, S1 gives cooking advice to S2 on how to do the mixing to get the best results; see Figure 5. Then in line 12, S1 asks for a translation of 'mix' into French, which S2 provides (line 13). In the episode above, we can see users following the instructions provided by the system, engaging with both the linguistic and culinary levels of the task, and providing help to their partner.

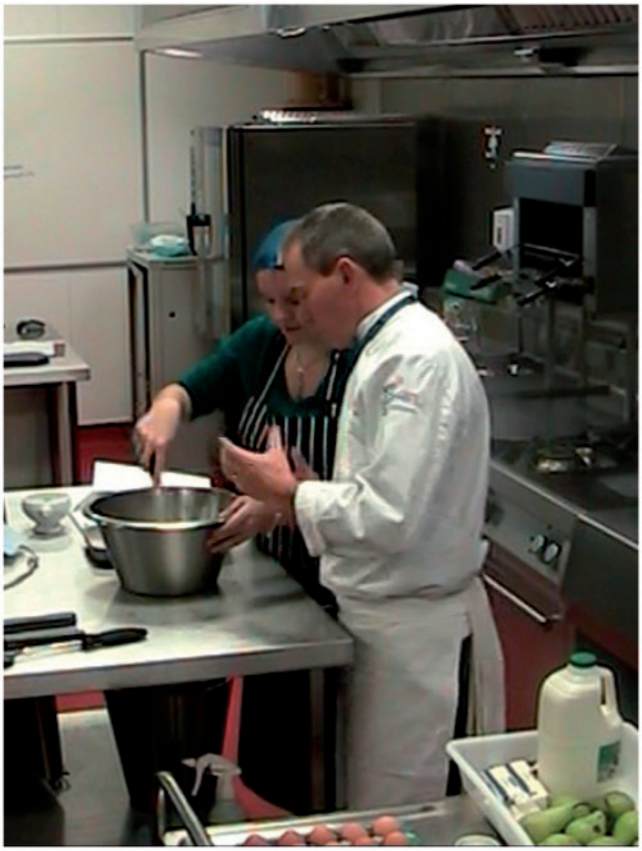


Figure 5: Mixing the ingredients (See Extract 1)

We can see evidence of incidental focus on form in lines 12–14, as well as of the information gap and information transfer processes targeted by TBLT. More detailed analyses of Digital Kitchen interaction and learning processes can be found in Seedhouse *et al.* (2013).

Iterative design: research

In this section, we describe the research element of the project, which had a dual function: first, to provide data for iterative system re-design, and secondly, to tackle the problem identified at the start. As mentioned above, the research design had to portray the process of users engaging with the digital environment as well as revealing the products of the learning experience.

The aim was to produce a real-word digital language learning environment where learners could simultaneously learn both French and how to cook a French dish, a linguistic and a non-linguistic skill. We investigated whether any aspects of French language were acquired by users of the kitchen by analysing transcribed video and audio data with their transcripts, post-test and interview data. The task targeted specific vocabulary items and these were tested via the digital display afterwards. In a post-task interview, we asked learners to identify any L2 words that they had learnt. So we were able to check for learning of specific lexical items by combining data from the interaction with post-test and report data. In TBLT, the accomplishment of a (non-linguistic) task is of key value, and we checked that participants were able to successfully cook meals as planned, by both observation and eating the dishes!

The data collection sessions lasted 60–90 min and we paired participants so that one was more skilled at French and the other at cooking. According to TBLT principles, this might create an information gap, thus promoting information transfer between the partners. In all, 36 audio and video-recorded sessions of paired adult learners (totalling 72) cooking in the kitchen were carried out over a period of 8 weeks. French levels ranged from advanced to absolute beginners. Many participants were British undergraduates studying French and other subjects, while others were college students of catering.

As noted above, we wished to portray the process of learners engaging with the digital learning environment in great detail. We therefore made audio and video recordings of users working on cooking tasks in all task phases. Each learner wore a microphone, and two cameras recorded task-related actions of the learners. These data were analysed using CA, a multi-disciplinary methodology for the analysis of naturally occurring spoken interaction that is now applied in a very wide range of professional and academic areas. There were two reasons for analysing the interaction between kitchen users in such detail. First, to provide the evidence of the learning process of French and catering skills, as illustrated in Extract 1. Secondly, the data fed into the process of iterative re-design. There are many components of the digital learning environment. We altered the configuration of components such as the language of instructions, timing of prompts and help, location of sensors, and how learners were paired. We then investigated the consequences of the changes for behaviour and interaction. The interactional and behavioural micro-detail enabled us to track the results of our configurational choices. We also collected digital sensor data using tracking hardware and software and these provided records of user actions, which fed into the system re-design process. In addition, we needed to establish the user's attitudes to the learning environment, given that we were trying to tackle a problem of motivation. There was therefore a post-task self-report interview and written questionnaire. Learners were asked what exactly they had learnt and were able to evaluate the experience. The questionnaire also asked about the learning supports they preferred (repetition, translations, partners, labels) and problems encountered; all of this fed into the re-design process.

As a result of the data collected, we improved the system design in several ways. For example, we noticed that some students over-used the translation

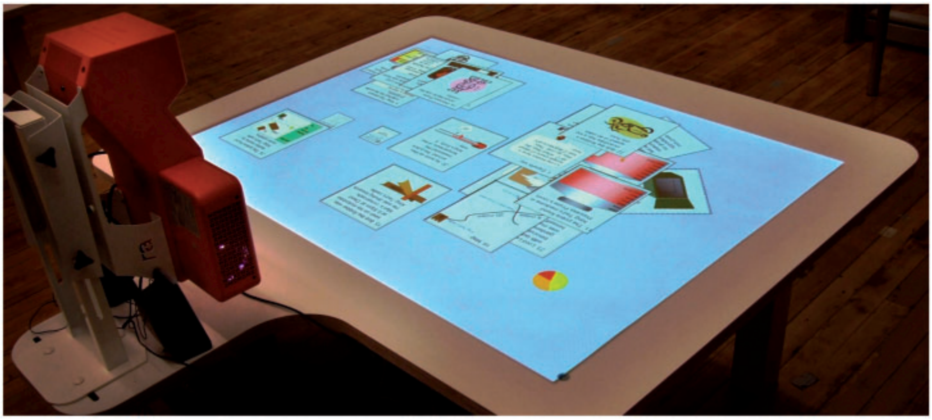


Figure 6: Digital tabletop

function. Therefore, we introduced a help function in the next iteration that showed photographs of items and videos of processes together with audio files and writing in the L2 instead.

TASK-BASED INTERACTION PROJECT

In this section, we describe another project that also involved applying digital sensor technology and TBLT to create a different digital learning environment. In this project, we again followed our model: Technology–Problem–Iterative Development and Research. The argument is that task-based interaction must be captured multimodally to be analysed. Moreover, as technology becomes more sophisticated and human interaction becomes more mediated by technology, the more the interaction assumes indexical qualities that require multimodal capture.

Technology

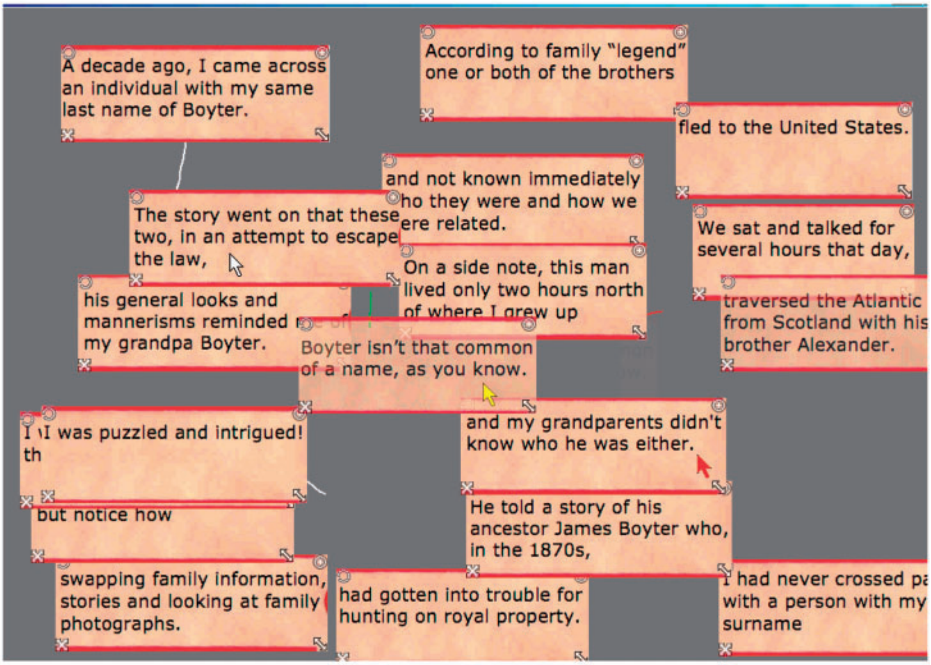
For this project, the technology with which we familiarised ourselves was digital tabletop (http://di.ncl.ac.uk/ilablearn/?page_id=20). Currently, a number of digital tabletops are available commercially, but at the time of the project, we learnt to use a prototype tabletop that was built by Culture Lab in Newcastle University (Figure 6).

Digital tabletops are multi-user, multi-touch interactive digital tables that combine interaction between users with the full use of digital media. They allow the development of collaborative, co-located educational applications and permit innovative task design. The table surface can be augmented with specific software designed to assist in a specific learning task. Participants can interact directly with the tabletop applications, allowing a substantial amount of data logging, which can be used for analysis of participant behaviour. This type of horizontal tabletop display can therefore adapt well to TBLT in terms of having groups of students working on a task using a shared space. Text, audio, video, and physical materials can be used on these tabletop displays and can be implemented in an interactive way. What gives this technology an advantage is its ability to enable groups of students to interact and collaborate on a task in a single space and keep an automatic record of who has done what when. The tabletops that we used can accept input from and track three participants using stylus pens simultaneously. Each stylus pen has a different colour, and the table's digital sensors can sense which pen is doing what at any given point in time and so it keeps track via video surface capture of how the participants are carrying out the task.

Problem

Learning the potential of the digital tabletop technology led to the realisation that it could be part of the solution to a specific problem, namely, how to analyse task-based interaction for language learning. Seedhouse (2004) pointed out that this has previously proved very difficult to analyse (particularly with convergent tasks such as information gap) because of its highly indexical nature and tendency towards linguistic minimalisation. Task-related actions and non-verbal communication could not be related easily to talk; even with video, it is difficult to distinguish a task-relevant action from any other action. The nature of convergent tasks tends to constrain the kinds of linguistic forms used in the learners' turns, and there is a general tendency to minimising linguistic forms. This is an example of what Duff (1986: 167) calls 'topic comment constructions without syntacticized verbal elements', which are quite common in task-oriented interaction. There is a general tendency to ellipsis, to minimise the volume of language used, and to produce only that which is necessary to accomplish the task. Turns tend to be relatively short with simple syntactic constructions (Duff 1986: 167). What we also often find in practice in task-oriented interaction is a tendency to produce very indexical or context-bound interaction, that is, it is inexplicit and hence obscure to anybody reading the extracts without knowledge of the task in which the participants were engaged.

Task-oriented interaction often seems very unimpressive when read in a transcript because of these tendencies to indexicality and minimalisation. L1 speakers engaged in convergent tasks in the world outside the classroom also often display some tendency towards minimalisation. However, this may give an unfair impression of task-based interaction, in that the full context of task-completion actions and non-verbal communication is not included. It is important to be able to analyse task-based interaction because TBLT (and interactionist approaches) claims that the interaction generated by tasks promotes L2 acquisition. However, if we were able to find a method of portraying all aspects of the interactional/pedagogical experience of task-based interaction from the learners' perspective, it may be that its value would become clear.



A decade ago, I came across an individual with my same last name of Boyter.

According to family "legend" one or both of the brothers

fled to the United States.

and not known immediately ho they were and how we ere related.

The story went on that these two, in an attempt to escape the law,

We sat and talked for several hours that day,

On a side note, this man lived only two hours north of where I grew up

his general looks and mannerisms reminded m my grandpa Boyter.

traversed the Atlantic from Scotland with his brother Alexander.

Boyter isn't that common of a name, as you know.

I I was puzzled and intrigued! th

and my grandparents didn't know who he was either.

but notice how

He told a story of his ancestor James Boyter who, in the 1870s,

swapping family information, stories and looking at family photographs.

had gotten into trouble for hunting on royal property.

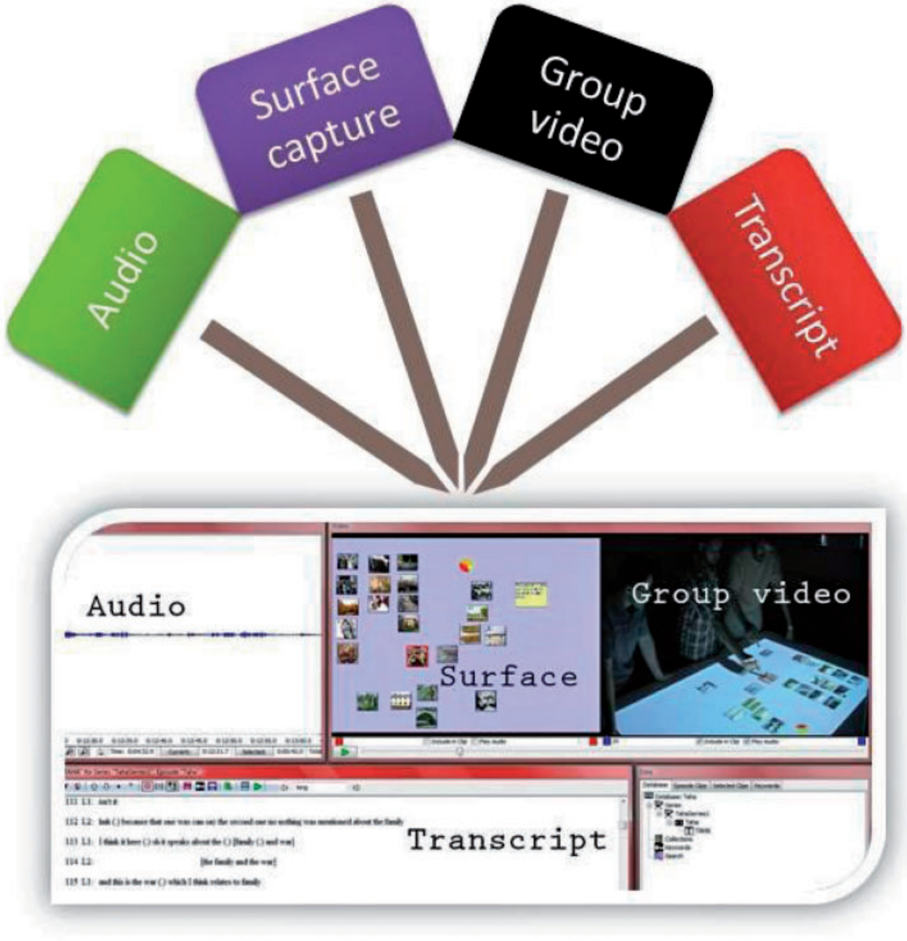
I had never crossed pa with a person with my surname

Figure 7: Surface capture of the jumbled text task used in this study

Iterative design

This section explains how Seedhouse and Almutairi (2009) developed a design using a combination of technologies to relate non-verbal communication and performance of the task to the details of the talk. They combined task-tracking hardware and software (digital tabletop), video/audio recording, and transcription. This enabled multimodal capture and a holistic approach, that is, one in which all elements of behaviour could be integrated in analysis. The task in this study was jumbled sentences text; a typical ELT classroom task that aimed to generate an information gap and interaction between students. The story was taken from a textbook designed for advanced learners of English. The story was digitised and embedded in the table (Figure 7). The software mixes up the pieces randomly on the tabletop, where the learners can manipulate them. They can move, rotate, and maximise the pieces of text using the coloured pens. The learners discussed how to re-form the narrative in the correct order. In terms of the distinction between convergent and divergent tasks (Duff 1986), this is a convergent task, as the learners need to agree on the order of the pieces as they rebuild the story.

Three people use three stylus pens, with different colours, and the tabletop records which pen does what when and hence, who has done what. In Figure 7, the digital surface recording shows which of the participants has



111 L1: isn't it

112 L2: huh () because that one was can say the second one no nothing was mentioned about the family

113 L1: I think it here () she speaks about the () [family () and war]

114 L2: [the family and the war]

115 L1: and this is the war () which I think relates to family

Figure 8: The different sources of data

moved the text where; the pens appear as arrows of different colours.[3] The specific advantage provided is that we have a separate record of task-relevant actions via the surface, which can be compared with other verbal and non-verbal actions from other sources. The surface of the tabletop works like a giant computer screen, so it is possible to use the screen capture feature that is built into the Windows operating system to give a video account of what takes place on the surface of the tabletop.

We were able to connect these data to data from the two pieces of video/audio recording equipment around the tabletop, as shown in Figure 8. The participants in this study were postgraduate international students and their English proficiency as shown by IELTS scores was quite advanced, the average

IELTS score for the participants being 6.5. Approximately 11 hours of audio and video data were recorded. The transcripts were produced from the audio/video recordings on the group video. After recording all the groups, the audio data were transcribed using CA conventions, then the transcripts were fed into Transana and synchronised with the relevant video recording.

Figure 8 shows how we were able to combine four data sources simultaneously for multimodal analysis, namely, audio, tabletop surface capture, group video, and transcripts. These sources of data were combined and synchronised using Transana software. When Transana starts, the screen (Figure 8) displays the two video sources, audio, and transcripts simultaneously. This presentation gives the analyst the convenience of examining all elements of task-based interaction as many times as needed. Moreover, the ability to review talk, non-verbal elements, and task-completion actions simultaneously enables analysis of the interdependence of these three elements, as we demonstrate below. It is proposed that task-based interaction can only be analysed adequately in conjunction with these two other elements.

Research: analysing task-based interaction using multimodal analysis

The iterative development of technology was intended to enable research, namely, of task-based interaction, thus solving the problem previously identified. This section explains how the research was conducted. As in the previous project, CA was used. It is appropriate, as it is able to integrate all elements of verbal and non-verbal interaction in a holistic approach, and has previously been used to analyse task-based interaction (Seedhouse 2004). We noted above that task-based interaction can be heavily indexical and it is therefore difficult or even impossible to read and analyse transcripts of talk without knowing what the learners are physically doing. The following transcript shows a heavily indexical encounter:

Extract 2

11	C:	and looking at family photographs	
12		(1.0)	
13	A:	and not (0.7) oh that's full stop (0.6)	
14	C:	yeh (0.8)	
15	A:	no (0.6) ok (.)	[photographs]
16	C:		[what is this]
17		(2.8)	
18	A:	ok	
19		(5.4)	
20 →	A:	the story went	

Extract 2

21		(1.3)
22	Y:	the story went on yeh yeh yeh (.)
23	A:	huh (.) do you [think]
24	C:	[I think] so yeh why not
25	A:	ok go ahead
26	Y:	*((moves piece 5 to her left))*

If we only had the transcript, one might think that A and Y were telling a story (lines 20 and 22) that would go on and some events would follow. A different picture emerges when we use the video view and surface capture that accompany the transcript. Seedhouse and Almutairi (2009: 324–6) carried out a multimodal analysis of the meaning conveyed by A in line 20 and how this meaning is understood by Y and C. In line 20, Speaker A places his pen on a piece of text (piece 5) that reads 'the story went...', gazes at it, and reads the first three words of the text out. A is thereby proposing that piece 5 should be the next piece of text in the story and in the following lines, the other students agree with this proposal. So an adequate analysis of this extract of task-based interaction is enabled by this system of multimodal capture. If one combines data from verbal elements, non-verbal elements, and task-completion actions, it is indeed possible to render the interaction comprehensible and analyse it as a multimodal speech exchange system. The multimodal approach is able to make visible the 'hidden' richness and complexity of task-based interaction, which is not evident from transcripts alone. Verbal elements can only be understood as carefully designed components of a multimodal system of communication. Previous studies of interaction in TBLT classrooms have confined themselves to verbal elements, and this study suggests that TBLT research has therefore not been able to reveal the full contribution of task-based interaction to the learning process. In the next section, we consider in more detail what is involved in the multimodal capture of interaction.

Speech, gesture, and multimodal corpora

To better equip corpus linguists with the means for examining patterns of gesture-in-use (i.e. language 'beyond the word' – see Knight 2011a), a surge in the development of multimodal corpora is being witnessed in the current research landscape (Reder *et al.* 2003; Ashby *et al.* 2005; Knight *et al.* 2006). Multimodal corpora are 'annotated collections of coordinated content on communication channels including speech, gaze, hand gesture and body language' (Foster and Oberlander 2007: 307–8), and are emerging as an invaluable tool for the study of communication. The specific methods and approaches that can be used to construct and analyse such data sets (particularly with a focus on gesture-in-use) are, however, something that is still very much in development.

The innovative contribution of the remainder of this paper is to question the potential for repurposing the sensors used in the Digital Kitchen study as a means of tackling a different problem/research question: the potential for using these devices as a means of capturing, encoding, and analysing patterns of language and gesture-in-use for the integration in multimodal corpora.

Compiling corpora and coding gestures

At present, no formally agreed, standardised approach exists for *recording* multimodal data sets for corpus linguistic study and, although each corpus tends to utilise a range of highly specialised equipment in a fixed, predefined, thus replicable, recording set-up, the exact nature of this setting is not necessarily consistent from one to the next (see Knight 2011b for further discussions on the current state of play in, and future directions of, multimodal corpus development). Different research aims and objectives will warrant different techniques for recording, depending on factors such as the scale/number of those under study, the artifacts or physical objects involved, and whether the recording is situated or not. The use of the tabletop devices does enable researchers to capture interaction with and interaction at specific points around a room, but such a system is not, for example, relevant for capturing episodes of interaction in more fluid and dynamic contexts (e.g. when outside or moving around a room and so on).

Extensive deliberation also exists about what aspects should actually be *marked up* and how; so which specific non-verbal behaviours (patterns of gesticulation) or prosodic features should be annotated and so on, to facilitate the analysis phase of the research. Coding schemes supporting the identification, representation, and analysis of different elements, components, and units that exist in spoken discourse proliferate, but there is a lack of such for marking-up non-verbal elements or for accurately integrating it with verbal elements.

The standard approach to mark-up and coding is generally manually driven, so is highly labour-intensive (Fanelli *et al.* 2010: 70) and often subjective. Given the advancing sophistication of digital devices, however, an increasing amount of (semi)automated coding tools are being developed in other disciplines such as psychology, computer science, biological sciences, and sign language studies, among others, which will potentially help with the standardisation, accuracy, and verifiability of the procedures used when constructing and encoding multimodal corpora for applied linguistic research.

Movement detection: automating the process

An early wearable form of such a tracker is the 'Dataglove' (created in 1987). Dataglove is a glove that features fibre optic cables that run down the back of each hand, each with a small crack in it. A light is shone through the cables so that when the fingers bend, light goes out of the cracks and by measuring light, hand poses can be identified. Revised versions of this glove have been developed, reporting an impressive 95% accuracy in detecting movements. The disadvantages of this system are, however, that it is both expensive and cumbersome and is only usable in fixed contexts, generally in lab conditions.

An alternative 3D-based method for capturing a range of bodily movements (beyond the head) and sequences of gestures was used by researchers creating the D64 Corpus (Campbell 2009). This corpus contains data from 4 to 5 people recorded over two four-hour sessions across a period of two days. The content is described as being 'spontaneous' and taken from a domestic environment, so is recorded in the home of one of the researchers involved in the project. Seven video cameras were positioned around the house along with two 360-degree cameras. A range of head-mounted and lapel microphones were also affixed to each participant along with four reflective sticky markers (also see Battersby and Healy 2010), three on the head, one on each elbow and shoulder, and one on the sternum (they were tracked by six OptiTrack cameras).

The use of the sticky markers provides the means for capturing bodily movements and sequences of gestures accurately, although they are 'not only time-consuming and uncomfortable' to wear but 'can also significantly change the pattern of motion' enacted by participants (Fanelli *et al.* 2010: 70). Their use also demands that researchers have access to what is very highly technological and expensive equipment, making it inaccessible to many.

Fanelli *et al. (*2010: 71) suggest the utility of alternative non-invasive 3D-capture techniques for gesture tracking and an alternative to sticky markers (focusing on patterns of head). For this study, participants were required to sit in front of a 3D scanner in an anechoic room and respond to a video that was played to them. The scanner provides detailed and reliable information about the position and movement of the head and its relationship with the spoken output from the speakers. However, as with the D64 system, not only is this approach particularly expensive (making accessibility an issue), the context in which the data are recorded is highly artificial, with speech being scripted, making it far removed from real-life naturalistic environments.

Similar, situated systems for gesture tracking that do not require participants to wear devices include the PlayStation Kinect and the Microsoft Move. These detect simple hand gesture signals via the use of sensors in webcams or motion controllers positioned in front of a user. Lochtefeld *et al.* (2011), for example, has started to examine the potential for using the Kinect device in the examination of non-verbal behaviour by tracking gestures used in sales conversations at a meat counter in a supermarket. Such devices are affordable and accessible to all, though, again, they are still somewhat limited insofar as they are required to be fixed in a specific location and cover only a small field of view.

It is clear that there is a paucity in the availability of tools that are inexpensive, accessible, and allow users to accurately track and recognise patterns of gesture use in dynamic contexts for linguistic research. Using the accelerometer device that is present in the Digital Kitchen sensors, the Culture Lab at

Newcastle University have developed innovative WAMs (wearable acoustic meters) that are small wrist-worn devices comprising an audio recorder and a tri-axial accelerometer (mapping up-down and left-right movements), which aim to do just this. They are worn like watches on both arms and can record for 48 hours without recharging.

The remainder of this paper reports on the preliminary piloting phase of using these devices. It discusses the processes by which the devices are being tested as a means for enabling the future construction of multimodal corpora.

DATA AND APPROACH

To support the piloting phase, data have been recorded in fixed contexts in the first instance. This is to enable episodes to be captured with video, to provide a key point of reference to which the tracking output can be compared. Ultimately, once trained, WAMs aim to help us to reproduce rather than film movement, so to capture forms of movement data 'in the wild'. This could give them the potential to be used to examine, for example, how gesture is used as a means of holding the floor; the relationship between gesture use, speaker incipiency, and interruptions; and how discourse-related gesticulation varies according to social and relational contexts.

Data were initially recorded in academic lectures, conference presentations (with a single or pair of speakers positioned at the front of a lecture theatre or seminar room), and research meetings. In total, 10 participants were recorded (4 females, 6 males) speaking for 20–65 minutes each, while wearing a WAM on each wrist. Participants were not told what to do, what to say, how to move, or what to wear for the recordings (long sleeves were permissible, for example), as we aimed to source data that were as reliable and naturalistic as possible (although the success of this is somewhat questionable owing to the existence of the camera).

A broad transcription for the audio output from the WAMs has been produced for each of the recordings, using ELAN.[1] Speaker tags have not been included, as individual tiers are instead used to differentiate one speaker from the next. To attempt to synchronise transcripts with the sensor output, transcript annotations are added (according to the time at which they begin and approximate duration) at the level of a turn, so are hinged around pauses in the monologue/dialogue, and attributed to points where a shift in topic appears to occur or at turn transition points. As a more fine-grained level of analysis, specific gesture movement sequences can be tagged with specific words and phrases that co-occur with them. This will allow for a closer analysis of the relationship between language and gesture-in-talk. While this is arguably not the most effective way of time-aligning speech within ELAN (a word-by-word level of alignment may be more appropriate), it is sufficient for the present study, which aims to test the feasibility of utilising the WAMs as a primary goal.

Once transcribed, the data are encoded with instances where gestures (defined here as non-verbal, expressive movements that play an integral part in determining *meaning* in discourse) enacted by the right, left, and both hands being marked up (using the video as a reference point). A second parse of manual encoding involved labelling the semiotic categories of the gestures, loosely based on McNeill (1985, 1992):

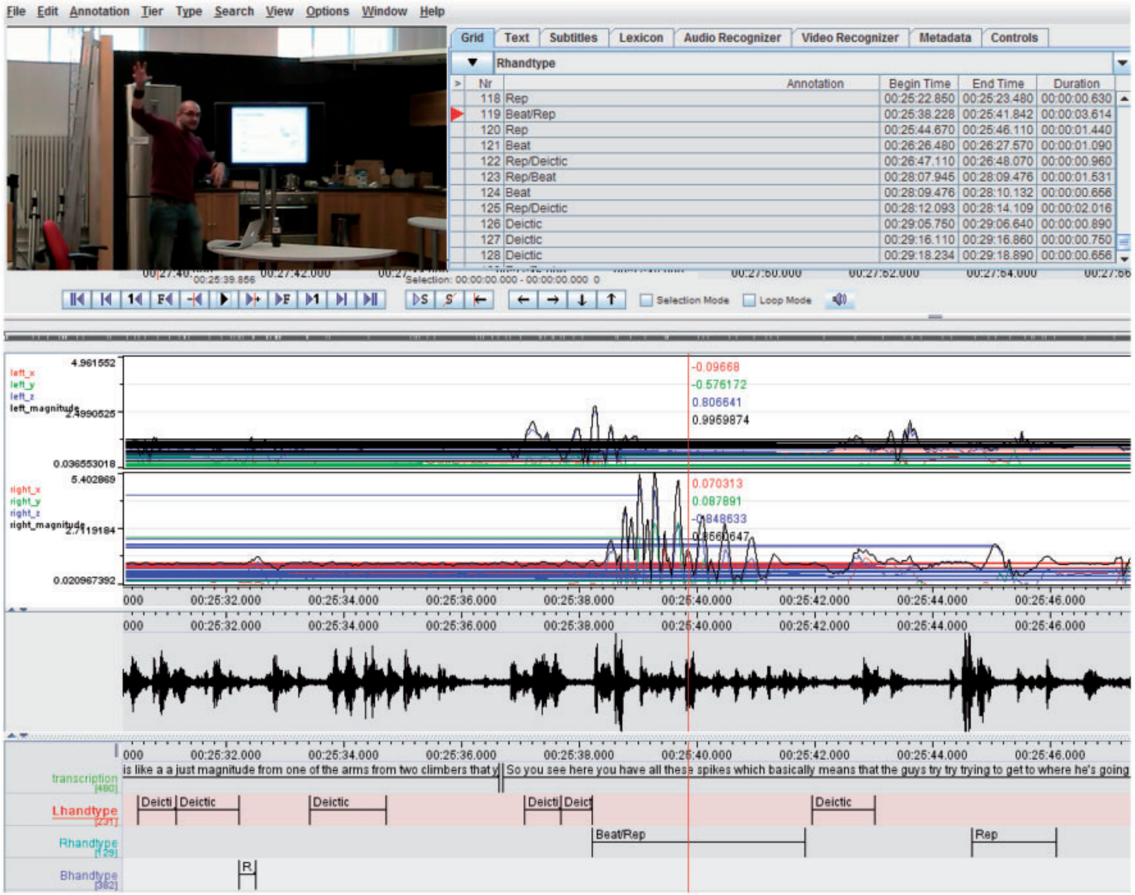
- *Iconics*: least consciously produced, but most culturally and contextually bound forms of gestures, commonly relating to concrete aspects of the scene.
- *Metaphorics*: relating to the semantic content but often used in parallel to sentences with abstract meanings (often combined with iconics, as is the case in the current study, also see Gullberg 2006).
- *Beats*: salient, repetitive movements (e.g. up-down/in-out motion) used to add emphasis and maintain the flow of conversation.
- *Cohesives*: repeated sequences of movements that help to create links in talk by providing visual relations to key semantic aspects of talk.
- *Deictics*: least semantically tied and most consciously produced forms, which are often pointing gestures used to signal to actual or abstract objects.

A screenshot from ELAN is presented in Figure 9.

The first two tiers of information (beneath the video) are the sensor's movement outputs from the left and right hand, followed by the audio output taken from the WAM. Beneath this is the transcribed speech from the episode followed by the coded semiotic gestures performed by the right and left hand, then both hands together. Each of these different tiers is synchronised by time and searchable via the facility in the top right corner of the figure.

To attempt to 'make sense' of this complex data set and to explore the potential utility of the WAMs for the aims specified above, the following semi-automated approach has been developed in collaboration with gesture recognition colleagues at the Culture Lab:

1. Distinguishing speech from silence in the audio output using a Voice Activity Detection (VAD) algorithm, to enable us to segment episodes where speech does and does not occur.
2. Distinguishing movement from non-movement based on the WAM data using a HMM classifier [Hidden Markov Model – this is a statistical algorithm, a (HMM) classifier, which automatically models time-series data].
3. Detecting episodes where speech and movement co-occur (mapping 2 on to 1).
4. Distinguishing gesture from behaviour (not every movement is a gesture)[2] – comparing the manually ascribed semiotic codes to the data outputted from 2.
5. Mapping episodes of gesture to specific acoustic patterns in the talk (based on 3 and 4).
6. Processing the semiotically coded gestures (using a hierarchical clustering algorithm) by pairing similar gestures in the data set together. This is



Nr	Annotation	Begin Time	End Time	Duration
118	Rep	00:25:22.850	00:25:23.480	00:00:00.630
119	Beat/Rep	00:25:38.228	00:25:41.842	00:00:03.614
120	Rep	00:25:44.670	00:25:46.110	00:00:01.440
121	Beat	00:26:26.480	00:26:27.570	00:00:01.090
122	Rep/Deictic	00:26:47.110	00:26:48.070	00:00:00.960
123	Rep/Beat	00:28:07.945	00:28:09.476	00:00:01.531
124	Beat	00:28:09.476	00:28:10.132	00:00:00.656
125	Rep/Deictic	00:28:12.093	00:28:14.109	00:00:02.016
126	Deictic	00:29:05.750	00:29:06.640	00:00:00.890
127	Deictic	00:29:16.110	00:29:16.860	00:00:00.750
128	Deictic	00:29:18.234	00:29:18.890	00:00:00.656

is like a a just magnitude from one of the arms from two climbers that y

So you see here you have all these spikes which basically means that the guys try try trying to get to where he's going

Figure 9: Coding and aligning the data within ELAN

useful for analysing the data in terms of how the instances relate to each other and to the episodes identified in 5.

The first two stages of this approach help to identify potential episodes for analysis, so the 'hot spots' in which gestures are likely to take place. Stage 3 provides further evidence to the incidence of *gesture* as, under McNeill's paradigm, *gestures* (proper) only occur in instances where speech is present. This in turn provides the basis for Stage 4. Stages 5 and 6 begin to identify specific acoustic and movement-properties of particular gestures, and their clustering according to their semiotic codes, to see whether interesting or common patterns emerge. The results gained from each stage will be used as a basis for developing a revised, refined approach to the semi-automatic detection and coding of gesture-in-use.

ANALYSIS

When *analysing* corpora, we typically preference the word or phrase as a way-in to the data. With this innovative data set, and by using this explorative approach, it may be more relevant to approach the data via the sensor output, semiotic codes, or acoustic properties of the speech as a way-in for identifying and exploring patterns within and across the different tiers of information. As this is 'in development', the testing of the appropriate way-in is an iterative process, with on-going adjustments and improvements to the approach being made (where necessary). Key methodological questions are also raised at every stage of the iterative process. For instance, questions over the appropriacy of using the McNeill system as a means of classifying the gestures have arisen (as this is not a fully taxonomic system), as has the perceived accuracy of the manually ascribed codes, in terms of the specific timing of the start and end point of a gesture (and the extent to which this can be mapped identically to the sensor output).

Adding further complexity to this is the fact that the potential meaning function of specific forms of gesture is highly variable and often difficult to interpret. While in spoken language, individual words (parts) are often combined to create sentences (the whole), with the individual parts determining the meaning of the whole, with gesture-in-talk, it is the complete gesture that determines the meaning of the individual parts (McNeill 1992: 19). They exist as 'complete expression[s] of meaning' unto themselves (McNeill 1992: 21), rather than being a sum of each of its individual parts. This means that individual gesture sequences can acquire various different and complex structures of meaning depending on their form, duration, attributed meaning function, and perceived synchronicity with on-going talk (in terms of mapping and meaning generation). The intricacies of this abstract and variable nature cast doubts on the extent to which such behaviours can be automatically tracked and defined in a reliable way (particularly as many of these behaviours can be largely idiosyncratic) – but this is something that is still being explored.

As the WAM devices only track up–down and left–right movements at present (a second prototype integrated with gyroscopes mapping is under development), the most promising successes to date have been in defining specific patterns of beat gestures (as these are the most salient form). The initial success rate for this is only around 60–70%, so there is significant scope for improvement.

FUTURE DIRECTIONS

While at present the results of this process are inconclusive (as testing is still underway), this section has outlined the potential for pushing the boundaries in applied linguistic research by utilising a current resource or tool (one which has proven 'worth') to address an altogether different research question or problem. Through future developments and further testing of this technology, it may be possible to take this research 'into the wild' and enable the linguist to explore patterns of gesture-in-talk in a variety of different discursive contexts. This will function to mark a step-change in research of this nature as it starts to provide us with the utilities for querying the notion of language use in 'context' (and the impact of context on linguistic choices), something that has long been noted as important in corpus enquiry, but never fully embraced or examined owing to the text-based nature of current corpus resources. The integration of the WAM devices when constructing corpora will, over time, enable us to record and integrate a wider range of semiotic resources in our linguistic 'data'. This will provide the impetus for generating richer descriptions of behaviour communication across a range of dynamic resources, contexts, and speakers, thus allowing us to examine the interactivity between the various modes and how they collaboratively create meaning. Our approach to, and perceptions of, using concealed, unobtrusive devices to capture such data will likely be challenged as a by-product of this future work, questioning current methodological conceptions of 'what is ethical?' in research 'in the wild'. It has initially been suggested that retaining only turn initial words and phrases from third parties in the construction of such data sets is ethically sound, as, arguably, individuals cannot be identified from such small amounts of language, so permission and informed consent is not strictly required. The legitimacy of this assumption, of course, needs to be questioned further. In relation to TBLT, the first two projects have shown that task-based interaction and learning can be researched outside the classroom setting. With digital sensor technology enabling multimodal tracking of how people perform real-world tasks on the move, the challenge will be to use the technology to form closer links between classroom and real-world tasks.

CONCLUSIONS

In this article, we have presented three projects that used our three-stage model for application of digital sensor technology, namely, Technology–Problem–Iterative Development and Research. In the first and second projects, a language learning environment was facilitated and tracked by digital sensor technology, while in the second and third projects, the technology enabled multimodal data collection and analysis. The research focus was therefore on how a digital learning environment might be designed, implemented, and evaluated. In all of the projects, the research focus has been on how to record and analyse the process of language learning through spoken interaction using digital sensor technology. This model is amenable to a variety of methodological approaches, as we have seen CA used in the first two projects and multimodal corpus linguistics in the third. There is currently an explosion in the number and range of new devices coming on the market that use digital sensor technology to track aspects of human behaviour. Many of these may be adapted for applied linguistic research, and the model and methods presented here may prove useful to researchers confronting the methodological challenges in this area in the future.

NOTES

[1] ELAN is a multimedia analysis and representation tool that is available for free online; see https://tla.mpi.nl/tools/tla-tools/elan/.

[2] Gestures are a specific form of movement/behaviour that can be interpreted as having a particular communicative function. When we walk down the street, we move our arms, but we are not gesturing. We are simply behaving. A gesture is dependent on the existence of the particular movement/behaviour in a communicative context and whereby it is either intended to (whether consciously or subconsciously) or interpreted as having a semiotic (meaning-related) function.

[3] The online version of this article displays colour versions of all figures.

ACKNOWLEDGEMENTS

Many thanks to project collaborators: Saad Almutairi, Patrick Olivier, Anne Preston, Dan Jackson, Philip Heslop, Thomas Plötz, Madeline Balaam & Saandia Ali.

FUNDING

The 'Language Learning in the Wild' project was funded by the Digital Economy Programme, a Research Councils UK cross council initiative led by EPSRC and contributed to by AHRC, ESRC, and MRC, grant EP/H043012/1. The 'Lancook' project was funded by the European Union Lifelong Learning Programme (Key Activity 2), grant 519076-LLP-1-2011-1-UK-KA2-KA2MP.

REFERENCES

Ashby, S., S. Bourban, J. Carletta, M. Flynn, M. Guillemot, T. Hain, J. Kadlec, V. Karaiskos, W. Kraaij, M. Kronenthal, G. Lathoud, M. Lincoln, A. Lisowska, I. McCowan, W. Post, D. Reidsma, and **P. Wellner**. 2005. 'The AMI meeting corpus' in Proceedings of Measuring Behavior 2005. Noldus Information Technology Limited, Wageningen, NL, pp. 4–8.

Battersby, S. A. and **P. G. T. Healy**. 2010. 'Using head movement to detect listener responses during multi-party dialogue' in Proceedings of the LREC Workshop on Multimodal Corpora. Mediterranean Conference Centre, Malta, 18 May 2010, pp. 11–5.

Campbell, N. 2009. 'Tools and resources for visualising conversational-speech interaction' in M. Kipp, J. C. Martin, P. Paggio and D. Heylen (eds): *Multimodal Corpora: From Models of Natural Interaction to Systems and Applications*. Springer, pp. 176–88.

Duff, P. 1986. 'Another look at interlanguage talk: Taking task to task' in R. Day (ed.): *Talking to Learn: Conversation in Second Language Acquisition*. Rowley.

Ellis, R. 2003. *Task-Based Language Learning and Teaching*. Oxford University Press.

Fanelli, G., J. Gall, H. Romsdorfer, T. Weise, and **L. Van Gool**. 2010. '3D vision technology for capturing multimodal corpora: chances and challenges' in Proceedings of the LREC Workshop on Multimodal Corpora. Mediterranean Conference Centre, Malta, 18 May 2010, pp. 70–3.

Foster, M. E. and **J. Oberlander**. 2007. 'Corpus-based generation of head and eyebrow motion for an embodied conversational agent,' *Language Resources and Evaluation* 41/3–4: 305–23.

Gullberg, M. 2006. 'Some reasons for studying gesture and second language acquisition Hommage à Adam Kendon,' *International Review of Applied Linguistics* 44: 103–24.

Knight, D., S. Bayoumi, S. Mills, A. Crabtree, S. Adolphs, T. Pridmore, and **R. Carter**. 2006. 'Beyond the text: Construction and analysis of multi-modal linguistic Corpora' in Proceedings of the 2nd International Conference on e-Social Science. Manchester, 28–30 June 2006.

Knight, D. 2011a. *Multimodality and Active Listenership: A Corpus Approach*. Bloomsbury.

Knight, D. 2011b. 'The future of multimodal corpora,' *Brazilian Journal of Applied Linguistics (BJAS)* 11/2: 391–416.

Lochtefeld, M., F. Dalber, S. Gehring, A. Krüger, and **J. Schöning**. 2011. 'Tracking pointing gestures to support sales conversations' in ACM International Conference on Human Factors in Computing Systems (CHI-11), Vancouver, BC, Canada, 7–12 May, 2011. pp. 1–4.

Long, M. 2015. *Second Language Acquisition and Task-Based Language Teaching*. Wiley-Blackwell.

McNeill, D. 1985. 'So you think gestures are nonverbal?' *Psychological Review* 92/3: 350–71.

McNeill, D. 1992. *Hand and Mind: What Gestures Reveal about Thought*. The University of Chicago Press.

Reder, S. R., K. Harris, and **K. Seztler**. 2003. 'The Multimedia Adult ESL Learner Corpus,' *TESOL Quarterly* 37/3: 546–57.

Seedhouse, P. 2004. *The Interactional Architecture of the Language Classroom: A Conversation Analysis Perspective*. Blackwell.

Seedhouse, P. (ed.). forthcoming. *Task-based Language Learning in an Immersive Digital Environment: The European Digital Kitchen*. Bloomsbury.

Seedhouse, P. and **S. Almutairi,** 2009. 'A Holistic approach to task-based interaction,' *International Journal of Applied Linguistics* 19/3: 311–38.

Seedhouse, P., A. Preston, P. Olivier, D. Jackson, P. Heslop, T. Plötz, M. Balaam, and **S. Ali**. 2013. 'The French digital kitchen: implementing task-based language teaching beyond the classroom,' *International Journal of Computer Assisted Language Learning and Teaching* 3/1: 50–72.

Skehan, P. 1998. *A Cognitive Approach to Language Learning*. Oxford University Press.