

Lecture 1

CALL, NLP, Corpus Linguistics

What is CALL?

CALL = Computer Assisted/Aided Language Learning.

For the purposes of this course we take CALL to embrace any computer software that is usable in some way to help language learners, whether intended for that purpose or not, and whether directly used by them, or used by someone else to create a conventional material (e.g. a coursebook) which learners use.

What is CALL?

Though the acronym "**CALL**" implies a limitation to language learning, we do not, as some do, distinguish that from computer aided language acquisition (**CASLA**). And we include in our scope language use by learners, and of course language teaching. Computer aided language testing (**CALT**) is often discussed separately from **CALL**, and for various reasons will not be much focused on in this course (lack of time and lack of the software!). We are also excluding use of computers in AL and ELT research in general (**CASLR**), and in the learning of linguistics rather than language (though there is an unclear borderline here, as much language teaching involves teaching about language, especially grammar, or raising awareness of language forms, and so resembles simple linguistics).

What is CALL?

There are many other acronyms and terms around with broader scope than **CALL**, or scope overlapping with **CALL**. They refer to areas of theory and research which have implications for CALL: e.g. **CAL**, **CAI**, **CBE**, **TELL**, **Telematics**, **HCI**, **AI**, **NLP**, **Corpus Linguistics**. On these neighbouring areas see Chapelle 2001 ch2 and Levy 1997 ch3 and pp77-82.

CALL 'tasks' include what may be otherwise referred to as games, exercises, activities, materials, even tests, and just 'ordinary use' of facilities like word processing. Sometimes they are fully determined by the program, sometimes they are largely in the hands of the teacher or learner using the software. They may be done in class or at home, etc.

What is CALL?

Thinking about **CALL** means thinking about many of the same things one considers when thinking about 'materials' for language learning/teaching (coursebooks, visual aids like posters or videos, pen and paper exercises, dictionaries etc.). Both involve something physical that teachers and learners use alongside a teaching method, syllabus etc. in a taught program OR which may be just used independently by the learner. Both have to be bought (or pirated). Both have a tangible form, but at the same time when exploited form part of a less tangible 'task' or the like. This parallel leads us to the conclusion that there are three main areas of concern (see Hubbard 1996 in ed. Pennington The Power of **CALL** for a fuller exposition, attempting to relate this to the Richards and Rodgers framework for analysing teaching methods):

What is CALL?

1) **Development/creation**. I.e. the principles and processes of writing software or authoring new materials within some existing software (Cf. Chapelle 2001 p166ff, and Levy 1997 ch4 onwards (esp. p104-108), for concepts rather than practicalities). Compare materials development, course book writing.

2) Use/implementation. I.e. how teachers use software with their learners (in or out of class, individually or in groups, for what sort of tasks, integrated with other aspects of the teaching-learning process or not, etc. etc.)... and how the learners use the software (which may be differently from how the teacher plans, or indeed entirely independently of school), their processes and strategies. Compare discussion of the role of materials like coursebooks or tapes in a course, different 'task types' they can be involved in, learner use of materials like dictionaries or cribs out of class unknown to the teacher etc... (Levy 1997 Ch4 onwards touches on ideas about Use repeatedly, esp p100-103; Jones and Fortescue ch14 old but practical)

3) Evaluation. I.e. how to decide what is good or bad software.... including inevitably considering what is a good or bad use of the software. Compare materials evaluation. (Chapelle 2001 Ch3).

HISTORY OF CALL

In terms of the development of hardware, program types, relation to ideas about language learning and teaching... This is filled out in class. See also Chapelle 2001 ch1 and Levy 1997 ch2 and the online <http://www.history-of-call.org/>

- The computer-as-big-as-a-room era. Entire courses like that of PLATO organised at a few universities. Audio-lingualism.
- The arrival of the home/school computer (Sinclair, Apple, BBC). CALL tasks as ancillary, and produced by many small publishers such as WIDA and even teacher enthusiasts. Attempts to fit it in with the Communicative approach.
- The era of the powerful PC (and Mac). Professionalisation of software writing but lack of transfer of much software from earlier platforms.
- PC + CD, multimedia. Software out of the hands of teachers, largely audio-lingual in mode. New attempts at entire courses.
- The era of the Internet. Teacher as selector. Learner-centred.
 - The future: convergence of media and 'omnimedia'
- Social networking?

Lecture 2

UUEG Software (Azar Interactive)

UUEG Software

<http://www.azarinteractiveonline.com/tour/>

Evaluation of UUEG

Before beginning the evaluation itself, it is necessary to give a brief description of the software, which is based on Betty Azar's book (2009). Due to space restriction, I will only provide an analysis of just one chapter of the book with intercepted description of the methods used in implementing the software in classroom. The analysed chapter is divided into four parts, each focusing on the following tenses: the present perfect, the present perfect progressive, the past perfect, and the past perfect progressive. Each section includes several quizzes, exercises and one crossword game, and these are followed by three main tasks covering listening, speaking and


reading comprehension (named by myself). To finish, there is a test that enables students to assess their achievements.

Evaluation of UUEG

Analytically speaking, the chapter follows Ur's framework (1988) for teaching grammar: presentation, explanation, practice, and test. The chapter starts with a preview of the tense, comparing it to, and/or contrasting it with, similar tenses – a method that is claimed to be effective by Walker (1967). Learners can either read or listen to the preview before examining a chart that exemplifies the tense. Following this, students are presented with a range of nearly all the typical mechanical drills, such as gap filling, error recognition, cloze, and multiple choices. Some of the quizzes come with animated pictures, and the exercises are represented in a linear progression – i.e. they become more difficult as the students advance. I would consider some of these exercises to be preparatory activities for the main tasks; for example, exercise 11 (Fig.1) prepares the students for the speaking task in exercise 16 (Fig.2).

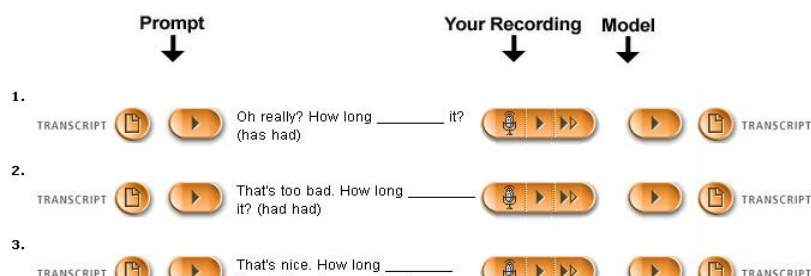
Within the program there are five main buttons located at the top of every page. These are made up of 'outline' (which outlines the whole chapter in detail), 'report' (enabling students to check their progress after each step), 'glossary', 'help' (where learners find help topics), and 'contents'.

Evaluation of UUEG



Below the cartoon, there are two columns of controls for a listening exercise:

	Your Recording	Model	
1.	I had not seen him in years.		TRANSCRIPT
2.	She had finished by 8:00.		TRANSCRIPT



Evaluation of UUEG

The listening task suggests that students listen to the recording of an international student's experience before answering the corresponding questions. A transcript of the dialogue is available.

In the speaking task (Fig.2) there is a 'record and compare' function that enables learners to listen to a prompt before reiterating the sentences whilst recording their speech. This enables them to compare their recordings to those of the model. Transcripts of the prompts and the model's words are available, and it is possible to play both of the recordings again and again.

The reading task comes in the form of a passage that includes some difficult hyperlinked words. By clicking on each, there appears a pop-up window that is linked to the glossary page. This displays the word's meaning along with a list of the other hyperlinked words, thus allowing students to check the meaning of other vocabulary. Multiple-choice comprehension questions follow the passage.

The above outlines what the software suggests for each task. However, it was I's decision to ask the students to discuss these undertakings in the specially-designed chat rooms, thereby making each task more communicative. I also decided to add further activities to each, and I discussed this idea later on in the evaluation. In order to motivate the students, I offered bonus marks for those who participate in the discussion and extra activities.

Chapelle (2001) evaluation scheme

For the purpose of this evaluation, it will be useful to begin with an outline of Chapelle's

Scheme (2001). Chapelle argues that CALL evaluation should be carried out using the theories of second language acquisition. There are two stages in her scheme: judgmental and empirical. In the judgmental stage, Chapelle (2001) analyses the software using two levels: the program and the teacher. In other words, she considers what learning conditions are set out by the software and what the teacher plans to do with the program respectively.

According to Chapelle (2001), however, this is not enough. She also addresses the question of what the learner actually does with the software by conducting an empirical evaluation. Whilst she focuses on different questions in each stage, she uses the same criteria in both. These criteria are: language learning potential, learner fit, meaning focus, positive impact, authenticity, and practicality. I shall judge the software by analysing the tasks using two of Chapelle's criteria: language learning potential, and learner fit.

Chapelle (2001) evaluation scheme

For the purpose of this evaluation, it will be useful to begin with an outline of Chapelle's

Scheme (2001). Chapelle argues that CALL evaluation should be carried out using the theories of second language acquisition. There are two stages in her scheme: judgmental and empirical. In the judgmental stage, Chapelle (2001) analyses the software using two levels: the program and the teacher. In other words, she considers what learning conditions are set out by the software and what the teacher plans to do with the program respectively.

According to Chapelle (2001), however, this is not enough. She also addresses the question of what the learner actually does with the software by conducting an empirical evaluation. Whilst she focuses on different questions in each stage, she uses the same criteria in both. These criteria are: language learning potential, learner fit, meaning focus, positive impact, authenticity, and practicality. I shall judge the software by analysing the tasks using two of Chapelle's criteria: language learning potential, and learner fit.

Lecture 3

CALL

Evaluation

CALL Evaluation

Basic definitions

'CALL software' here can involve any software or programs potentially usable by language learners in connection with learning/teaching or use of language (esp. EFL/ESL). That includes both material claimed as designed for this purpose ('dedicated'), and that not. The latter includes both specific programs like adventure games for native speaker children, and 'generic' or content free software like email or word processing. It also includes whatever hard copy support materials, booklet etc. any software comes with. See further our Intro.

"Evaluation is a matter of judging the fitness of something for a particular purpose" (Hutchinson and Waters 1989: 96). 'Evaluation' therefore implies an activity where something is declared suitable or not and consequent decisions are to be made or action taken. Evaluating something therefore is not the same as researching it, though research may be done to find out things which then inform the value judgment and hopefully make it better. Research on its own may just end up with information, not judgment and action.

CALL Evaluation

CALL software and general teaching materials and tasks - a parallel?

Much of what we say below about evaluation of CALL software is similar to what one would say for 'materials evaluation' generally in language teaching. CALL software is often analogous to an individual exercise or task in a book, though some series of CDROMs constitute entire courses and so are parallel with complete coursebooks. The parallel is valuable... up to a point. There are some important differences, however.

Firstly, a book is not typically dynamic or interactive; a program, by contrast, may not always present an exercise the same way every time you use it, and can usually give some response to the user dependent on what they click or type in. That is why CALL programs have often been seen as replacing a teacher rather than just teaching materials, though that clearly does not fit all software.

Secondly, a book is more limited in its media capability. CALL can involve sound as well as pictures, diagrams and text all in the same package.

Thirdly, use of written materials has few technological prerequisites: eyes and a desk to put them on will do. CALL by contrast requires computers, network access etc.

Fourthly, the language content of material in a coursebook is essentially unalterable, while some CALL software allows 'authoring': i.e. the teacher can put in his/her own choice of text, words etc. for the program to make an exercise out of, or whatever. In fact some software, such as a wordprocessing program, is essentially content-free and is nothing unless someone enters text to make an exercise, or designates a task for learners to do with it (see next).

Fifthly, the activities to be done with each section of a coursebook are usually heavily constrained by the book itself, though there may be some latitude for the teacher to implement exercises in different ways, and of course skip some material. A CALL program on the other hand may be very constrained (e.g. a hangman game), or may be almost entirely open in this respect (e.g. email). The last two are important for evaluation, as they make it hard to draw a line sometimes between evaluating the software and evaluating the specific language material a teacher has put in, or a specific task done with the software which is not determined by the software itself. I.e. the borderline between evaluating software 'in itself' as a material and evaluating some proposed or imagined use of the software becomes impossible to maintain.

CALL Evaluation

The importance of evaluation

Evaluation is one of three key aspects of CALL that need consideration: Creation, Use and Evaluation.

CALL shares one important thing with teaching materials and tasks in general. All these are under-evaluated. Just as new coursebooks and types of task are constantly being proposed and promoted by their creators ... and adopted and used... so are CALL programs and activities (Chapelle top of p10). What rarely happens is any proper evaluation of the value or effectiveness of any of this.... by teachers or researchers. Correction: some teachers may well do a lot of evaluation of what they use... but, if so, it remains within their personal teaching process and is not published. Hence we have no idea how much of this goes on, or what evaluation methods and criteria are used; furthermore, nobody else gets the benefit of the information arising from the evaluation.

CALL Evaluation

The three key components in CALL evaluation

Mostly evaluation cannot be done in the abstract. I.e. things are rarely universally good or bad. With CALL you may feel some programs have features which in NO situation would be any good. Possible candidates for 'universal' status could be software glitches (e.g. the program crashes whenever the help icon is clicked) and inaccuracy of language (e.g. multiple choice exercises where the option counted as correct is actually wrong). However, a lot is really 'relative' and it is as well to start off thinking of everything as potentially relative than the reverse. As Chapelle says (2001 p52): 'Evaluation of CALL is a situation-specific argument'.

Clearly most features may be good for one type of person, situation etc. but bad for another. For example the kind of vocabulary included, the kind of computer knowledge required to work it. This is as true of general materials evaluation as of evaluation of CALL specifically. So one important aspect of evaluation is to establish the specific users (learners and teachers), situation, purpose etc. etc. that you are evaluating the materials for. This means that you cannot really evaluate without also thinking of how the material will be used in the learning and teaching process. It is quite possible for one and the same program to seem 'good' when used one way with a class and 'bad' used another way, or with a different class. Software and materials evaluation in ELT, then, can be seen as an activity where you match materials to teaching/learning situations. I.e. there are three things to think about -

- (a) the nature of the materials/software: describe in detail what it consists of/does (especially if your account may be read by someone not familiar with the program). As mentioned above, this may extend to analysing the specific task it is used for/in. 'It's not so much the program, more what you do with it' Jones 1986.
- (b) the nature of the T/L situation, the learners and their needs, uses etc.: describe in detail (not just 'intermediate learners'). Levy 1997 has several somewhat theoretical sections on describing CALL e.g. p108f, 156f, 173f.
- (c) a rating or judgement to make of suitability of one of the above for the other, with due attention to relevant universal principles of good teaching/learning; explain how this is going to be done (e.g. introspectively or empirically - see below) and execute it.

CALL Evaluation

One may of course do that for just one piece of software at any one time, but it is often easier to evaluate two or more programs of the same type together.

Comparisons are often revealing. In addition, one may often usefully compare a CALL activity/program with a non-CALL (pen and paper) counterpart, as has widely been done in writing research (pen versus wordprocessor).

Furthermore you can deal with the above three components one of two ways round:

- (i) You can think of a specific type of learner, teaching situation, required activity etc. first and consider whether or not each of a set of materials/each separate activity in a software package would be suitable or not for that one case. A teacher in the field is likely to work this way ("Would this suit my class?"). It is certainly easier to produce a clearly focussed evaluation that way. Note: in this course the

idea is not just to evaluate CALL for ourselves as users, but to think further afield of some potential learner user type.

(ii) You can start with the materials/program and consider what range of people, situations, ways of being used etc. etc. it would suit and which not. The courseware 'reviewer' in a journal, and perhaps some of us here as AL/ELT people not currently teaching any learners directly, may prefer to think this way. When software comes with claims by its authors of what learners it is suited to, this can be a way to proceed. (But this can degenerate into letting what software is available drive what one does rather than the reverse Chapelpe p44)

CALL Evaluation

When the evaluation is done

It is also worth noting that there can be several types of occasion when evaluation of teaching materials, including CALL, may occur (overlooking evaluation done while the software is actually under development):

- 1) Evaluation of materials **prior** to purchasing them or creating access to them for any learners. I.e. as a result of evaluating materials you decide whether to buy or adopt them or not, for some specific learners. (Direction i usually, though ii is also possible).
- 2) Evaluation **after** purchase or otherwise acquiring availability of software, but before use. Here usually the question is what learners it would suit. So the consequent action is to use it with/recommend it to these learners not those, and so on. (Direction ii, or i).
- 3) Evaluation **after** the program has been **acquired** and **used** with some learners for a bit. Here the question is whether it was a success and the action is to use/not use the program again with these or other learners, or to alter the way it is used in some way. (Direction ii).

This account is focused more on 1 and 2, since most of us are not teachers who have just been using CALL with any actual learners, but the same ideas pervade all three situations. In all of them you decide if the materials are good or bad, not just what they consist of or 'do' etc.

CALL Evaluation

Who evaluates

The evaluators we are thinking of here are primarily language teachers, though of course other people evaluate materials too - curriculum/program planners, government education departments, reviewers writing for journals, researchers in applied linguistics...etc. In the realm of CALL, it is especially necessary for teachers to be good at evaluating. There is a lot of poor material about; publishers are especially prone to hype; curriculum designers who might evaluate to choose suitable coursebooks for a course are less likely to extend this activity to CALL, so the job is left to the teacher; only a few teachers write their own CALL software (compared with the number who might write bits and pieces of their own non-CALL teaching materials) - most rely on professional products (though remember programs may require or allow some teacher 'authoring').

The judgmental evaluation

Methods of evaluation (B): Empirical evaluation

Other methods of evaluation generally require much more work, and for the materials to have been used for some time by learners/in actual classes (compare situation 3), so they are often firmly fixed in a specific teaching/learning situation (b). However, they do move away from the purely introspective approach. These are the ones that incorporate activities that are just like those we would otherwise regard as typical of regular empirical 'research' - measurement, surveys etc. I.e. they may entail using questionnaires and interviews, systematically observing, eliciting 'think-aloud' data from software users, or testing users. They may mean doing 'studies' (experimental or not) comparing the success of one material against another and so forth, or indeed doing 'action research' with CALL. (See Chapelle, Jamieson and Park 1996 in ed. Pennington *The Power of CALL* for an overview of types of empirical research done on CALL classified by the kinds of methods used; and Chapelle 2001 pp66-94 for a more detailed coverage, in relation to CALL tasks of the more communicative type, and classic SLA research issues looked at in CALL)

In themselves these 'research' type activities are non-evaluative, in the sense considered here (except action research). They are best seen as scientific means of gathering facts and testing hypotheses which can then either remain as cold statements of fact about what the effectiveness of the materials is or what people's opinions about them are, or be exploited for practical ends as part of an evaluation exercise - i.e. to make decisions like those described at the start.

The judgmental evaluation

Examples are:

Doing a survey of teachers and/or learners who have used the material and finding out how they use it, their difficulties, attitudes to the interest and usefulness of the content, tasks etc. Checklists can come in here again. E.g. one can base a questionnaire to users around the same set of (a) and (b) points that might otherwise be the points one asks oneself about in A above.

Observing a class using the program, taping and making systematic notes on their difficulties, actions, strategies, what they say, the teacher's involvement etc. Or one can ask learners to keep a diary of their reactions.

Getting the computer to store records of actions performed by learners using a program and analysing them to infer learner strategies and processes. (E.g. revisions when wordprocessing, accesses made to an online glossary when reading). Example in T. Johns 1997 'Contexts' in ed Wichmann et al *Teaching and Language Corpora* (Longman).

The classic research comparison of those using one program with those using another differing in a small or large way (or no program... just doing non-

computer equivalent tasks) over a period, with before and after tests to check on how much has been learnt.

If A type and B type evaluation are both done, the connection between the two needs to be spelt out. If the A evaluation resulted in adoption of the software, did the B evaluation show that was a good decision?

أطيب الامنيات للجميع بالتوفيق والنجاح

Edit by Heaven Lover