ASp
la revue du GERAS

60 | 2011
Varia

From the cultural hegemony of English to online informal learning: Cluster frequency as an indicator of relevance in authentic documents

Geoffrey Sockett

Electronic version

URL: <http://journals.openedition.org/asp/2469>

DOI: 10.4000/asp.2469

ISSN: 2108-6354

Publisher

Groupe d'étude et de recherche en anglais de spécialité

Printed version

Date of publication: 1 November 2011

Number of pages: 5-20

ISSN: 1246-8185

Electronic reference

Geoffrey Sockett, « From the cultural hegemony of English to online informal learning: Cluster frequency as an indicator of relevance in authentic documents », *ASp* [Online], 60 | 2011, Online since 07 October 2014, connection on 04 November 2020. URL : <http://journals.openedition.org/asp/2469> ; DOI : <https://doi.org/10.4000/asp.2469>

This text was automatically generated on 4 November 2020.

Tous droits réservés

From the cultural hegemony of English to online informal learning: Cluster frequency as an indicator of relevance in authentic documents

Geoffrey Sockett

Introduction

- 1 In the field of language learning and teaching in France, researchers seek to underline the importance of linguistic diversity through the study and promotion of plurilingual skills. Specialists in the didactics of English should however be careful that this quest for linguistic pluralism does not mask the unique position occupied by the English language in France and the effects this position may have on learning. Swift and profound changes are taking place in the way English language media are used in France, most particularly online media. It is therefore important to continue to measure and analyse these changes in order to get a better grasp of the learner motivations so as to organise the formal teaching of English by making the best of them.
- 2 This article is part of a broader study on how French learners of English read, listen to and use the target language informally in a range of online activities. These phenomena will be referred to in this study as the Online Informal Learning of English (OILE). The aim in this instance is to seek to characterise the type of language which makes up the majority of informal listening activities engaged in by learners of English in their leisure time. This analysis should provide pointers as to the relevance of this input material for learning in general and more formal learning in particular.
- 3 After a brief review of the pervasive nature of English in France, the current state of research into OILE will be presented, along with the theoretical framework which underpins it. A research project will then be described in which a corpus of scripts from

television series frequently viewed in English by French students is analysed. The most frequently occurring clusters of 4-word groups in this corpus are ranked, following the methodology of Chen and Baker (2010), and these clusters are analysed so as to determine their relevance with respect to the types of language skills suggested by the action-oriented approach. The frequency of these clusters in everyday English is also established by correlating their ranking with that of the British National Corpus (BNC). Finally, the pedagogical implications of this study are outlined, and areas of further study in the field are suggested.

Background to the study

- 4 Antonio Gramsci's (1991) notion of cultural hegemony has been widely taught in Cultural Studies over the past forty years. It is clear that the influence of language and culture can be measured in terms of the impact it has on everyday life. In France as in many other countries in the world, the most frequently watched TV series are American productions such as *House* or one of the *CSI* franchises. Films and series which once had French titles such as *La Guerre des étoiles* or *Le Frelon vert* are now called *Star Wars* or *The Green Hornet*. With the expansion of the Internet since the mid-1990s has come an increasing exposure to websites such as YouTube, Facebook and Wikipedia, in which English is often used.
- 5 Within this growing field of influence, one dichotomy has remained in place, namely the distinction between countries where English language television programmes are subtitled and those where the programmes are dubbed into the local language. This distinction between countries with a large population such as Italy, Germany, France and Spain, and those whose smaller populations rendered dubbing uneconomic (Holland, Sweden, etc.), is generally considered to have had a profound effect on the levels of English listening comprehension skills in those countries (Ruperez Micola *et al.* 2009; Vanderplank 2010). Indeed Bonnet (2002) ranked French adolescents as the weakest in English listening comprehension from a group of eight Western European countries.
- 6 Our preliminary studies in this field (Toffoli & Sockett 2010; Sockett 2011) suggest that informal online activities such as downloading or streaming English-language television programmes may contribute to a change in this dichotomy, as viewers (particularly learners of English) switch from the television screen to the computer screen as their primary means of viewing programmes, and as they become dissatisfied with the experience of watching dubbed material once they have got used to the original versions of series. Hence increased exposure to authentic English is likely to be a reality for many learners and this change is likely to have a positive effect on listening comprehension skill levels overall.

The online informal learning of English

- 7 A growing realisation that learning takes place both inside and outside the classroom (Cross 2006) has led to an increasing interest in studies into lifelong learning and the relationship between formal and informal learning, particularly in the field of languages (Stevens 2010).

- 8 We have defined informal language learning as emerging from a communicative intention, which implies that it does not take place according to a set timetable or in the context of an organised learning activity and indeed that the learner is not necessarily aware that learning is taking place (Toffoli & Sockett 2010). This definition is in line with that of Tissot (2004) who considers that learning results from daily activities related to work, family or leisure, and the European Commission (discussed by Stevens 2010) which points out that these activities may not be recognised by the individuals themselves as contributing to their knowledge. Such a definition differs from that of Cross (2006), who sees informal learning as an intentional activity aimed at learning to do one's job through networking activities.
- 9 Just as the global reach of English has primarily made itself felt since the beginning of the Internet age, the informal learning of English in many western countries takes place primarily through online activities. Our initial quantitative study (Toffoli & Sockett 2010; Sockett 2011) established that a majority (60% of 225 students surveyed) of non-specialist learners of English (students whose major area of study was not English but who had one to two hours of English classes per week to fulfil a language requirement), downloaded or streamed films and television series in English from the Internet on a weekly or monthly basis. Most students also read information in English from a variety of websites. As many as a third of the 225 students surveyed reported regularly using English as a lingua franca to write to friends on social networking websites such as Facebook, this means of communication being cited more frequently than emailing, using audio or text chat services or participating in forums in English.
- 10 The preponderance of comprehension activities (listening and reading) over production activities (writing and speaking) has led to comparisons with the Input Hypothesis approaches to language learning of the 1970s and 1980s. In these approaches, Krashen (1982) and others suggested that a sufficient quantity of comprehensible input was a key to language learning. We have argued (Toffoli & Sockett 2010) that it is also useful to consider OILE within the framework of the Output Hypothesis (Swain 2005), Incidental Learning (Rieder 2003) and Dynamic Systems Theory (Van Geert 2008). Swain's Output Hypothesis sees "languaging" or engaging in communicative activities as a key to learning, while Rieder's work underlines the fact that, at more advanced levels, language learning often takes the form of incidental vocabulary acquisition in the course of other activities and is often not a deliberate activity.
- 11 In a follow-up qualitative study (Sockett & Toffoli, forthcoming), five students kept diaries of OILE activities over a period of two months. This provided insights into the amount of time spent on these activities, their regularity, and the way in which the affordances of the online interfaces facilitated language learning. Information about student perceptions of learning was also gleaned. It emerges that students spend some 20-25 hours per month involved in OILE activities, including some 4-8 hours of television viewing in English. They tend to visit the same websites repeatedly over a long period and engage in durable exchanges with other English users. The role of on-demand music services such as Deezer or Spotify was of particular interest in this study. Whereas English-language music was once simply heard on the radio, it is now more frequently chosen by the listener and listened to more actively. For example, the study revealed frequent references to a concurrent use of lyric websites such as lacoccinelle indicating that listeners are now trying to process songs for meaning, rather than just

treating them as background music. In the context of OILE activities, learning is most often experienced as vocabulary acquisition, and students are also aware of the progress they make in understanding the meaning of television series and films.

The theoretical framework of OILE

- 12 We have already stated that OILE is a largely incidental process which comes about as a by-product of a communicative intention. In previous publications (Toffoli & Sockett 2010; Sockett 2011) we have discussed the importance of Dynamic Systems Theory as a key to understanding how learning comes about in a non-linear manner as a result of complex interactions between many different elements in the learning system. The pervasive nature of the English language in France, which, for repertoires as diverse as scientific language, computer use and popular music, functions more as a second language than as a foreign language, leads to many such interactions between the learner, his or her context, formal learning material from English classes and informal contacts with authentic documents and other English users on the Internet. Indeed the *Académie française* considers that 5% of all words used in French today are of English origin.¹
- 13 A theory compatible with this view of learning is CREED (Ellis 2007), an associative cognitive view of language acquisition which argues that learning is Construction based, Rational, Exemplar driven, Emergent and Dialectic. This perspective emerged as a cohesive theory from Ellis' earlier work on areas such as Implicit learning (1994) and Frequency effects (2002).
- 14 Construction grammar (Legallois & François 2006) considers that constructions are the basic elements of language and that they are learned from their repeated usage in the speech community. These constructions may be groups of words which frequently occur together such as "as soon as" or even "and it is the" and are not necessarily idioms like "to strike while the iron is hot". The concept that language occurs in recurrent groups is not a new one. Indeed Sinclair refers to this phenomenon as "recurrent co-occurrences that a word has with its collocates" (1991: 17), in developing his idiom principle. Terms such as phraseology, formulaic sequences, clusters and recurrent word combinations have been used to describe these constructions, and their importance in fluency, as reliable chunks of language which reduce the processing burden on the speaker, has been broadly discussed in the literature. So fluent speech can be seen as a series of formulaic structures which require little grammatical processing and thus release working memory capacity to attend to the macro-structure of the utterances.
- 15 Learning in this model is rational to the extent that a combination of factors such as frequency and recency of exposure and salience² in the target material can be used to predict language development.
- 16 The argument that learning can be exemplar-driven rather than rule-driven depends on the existence of prototypes or predominant examples in the target material. Hence a learner will be frequently exposed to a structure in which the same verb is predominantly used and will learn the example, such as: "he gave it to me", before later attempting to extrapolate on the example by using other verbs in the same construction ("handed", "presented", etc.).

- 17 The emergence of patterns in a complex system refers to the fact that knowledge is self-organised during the learning process, rather like a pile of sand which goes through phases of accumulation and settling which are independent of changes in the input flow (Van Geert 2008). Hence, in language acquisition, learners experience spurts in fluency as they combine larger chunks of formulaic language which seem like rapid changes in their language skills even though the quantity of input they receive remains the same.
- 18 Finally the dialectic nature of learning underlines the tension between the current state of the learners' interlanguage and their perception of the language of the speech community around them. This tension can drive learning and, in the case of informal learning, is promoted by the high level of exposure to spoken and written English reported by informal learners.
- 19 The importance of the different elements of CREED for the present study is quite clear. Learners encounter a large number of constructions in the English they listen to on a regular basis and these can serve as exemplars in the emergent language system. The extent to which these constructions are frequent and representative of English as it is spoken in the real world will be examined in our research. It can be argued that television series viewed by learners for reasons of personal interest offer situations in which it is easy to infer meaning due to knowledge of the characters and the repetitive dramatic structures of situation, disruption and resolution which predominate. Therefore frequency and salience can be inferred and the learning process may be viewed as a rational consequence. This type of learning is dialectic in that learners notice the gap between their current understanding of dialogue in the series they are watching, and a "comfortable" level of comprehension (Sokkett & Toffoli, forthcoming). The fact that television viewing is just one of a range of activities engaged in by informal learners means that other opportunities exist, notably social networking activities, to send and receive messages and thus experience a language community which may enable them to notice the gap between their productions and the replies they receive.
- 20 Chapelle considers that the importance of CREED for understanding language learning lies in the relationship between repeated exposure to input and the development of proceduralised skills, stating that learning tasks should provide "optimal amounts and circumstances of exposure to the targeted linguistic areas to be learned" (2009: 747). It may be argued that learners create such tasks for themselves within the complex system that is informal online language learning.
- 21 In the field of neuroscience, a construct which may prove useful in understanding informal listening is the concept of additional sensory stimulation developed by Wright *et al.* (2010). In their study, subjects are trained to distinguish between musical tones. Between the periods of active training, a control group is left in silence while the experimental group hears music. The findings of this study are that the exposure to music during the breaks between training sessions has a significantly positive effect on the active learning process even though attention is not being paid to the music. An important finding of the research is that the shorter the gap between active and passive activities, the greater the impact, with a greatly reduced impact being observable in the case of a break of only four hours between active and passive learning phases.

- 22 In order to apply this research to the field of OILE, it is necessary to consider that listening to music and watching television series and films in English may be regarded as elements of additional sensory stimulation which positively affect learning even though attention is not necessarily being paid to content. This is particularly the case with English language music. The emergence of mobile multimedia devices such as MP3 players and similar functions integrated into smartphones increases the likelihood of learners being exposed to the target language both before and after active learning experiences such as university English classes. Sales figures for music downloads in France suggest that more than half of the albums downloaded are in English.³ Interactions between regular active viewing of English-language media and listening to English music could also constitute an example of additional sensory stimulation.

Methodology of this study

- 23 Within this theoretical framework, a number of questions remain as to the frequency, relevance and salience of constructions which learners may listen to in the course of their online leisure activities. The creation of a corpus representative of a typical year of viewing of English-language programmes by French students is a pertinent way to observe and characterise the contents these learners are exposed to. It may also provide insights into the implications of these activities for English teachers.
- 24 The methodology adopted for this study follows that of Chen and Baker (2010) who work in the field of academic writing. Their study entitled “Lexical bundles in L1 and L2 academic writing” compares the frequency of 4-word constructions in scientific texts written by L1 English academics, L1 English students and L2 students whose native language is Chinese.
- 25 For Chen and Baker, the choice of the 4-word cluster as the optimal block size to study relates to the manageable number of examples generated in a corpus of this size, and to the widespread use of 4-word blocks in other similar studies. Using WordSmithTools corpus analysis software, they establish a ranking of lexical bundles by frequency order for these three categories of authors from three corpora of about 150,000 words each. Identifying which constructions are over and under-represented in the Chinese productions as compared to the expert productions enables Chen and Baker to make suggestions as to how Chinese scientific productions may be made to sound more idiomatic.
- 26 The advantages of this approach for a study on the OILE field are numerous. The general approach of Chen and Baker is coherent with a CREED view of language learning, emphasising the importance of lexical bundles, in the target material. Ranking these constructions by frequency also allows the researcher to judge their relevance for the learner while a comparison with another corpus indicates the likelihood that the structures will be heard and used elsewhere. The clearly outlined methodology of the study also lends itself to replication with the widely available concordancing tool, Wordsmith 5.
- 27 In the present study, a corpus (given the name HHOLD, from the initial letters of the television series analysed) of material likely to be listened to by non-specialist students of English over a period of twelve months was developed. Our previous qualitative diary study of five students and quantitative study of 225 students provided

information as to the quantity of television series viewed by informal learners and the names of the series they were most likely to download or stream from the Internet. The students involved in these studies were undergraduates at the University of Strasbourg, majoring in Sociology, Fine Arts and other humanities. Their level in English had been tested as B1 and B2 according to the *Common European Framework of Reference for Languages* (CEFRL) scale by a placement test at the beginning of the year. The studies indicate that many non-specialist students watch around 70 hours of television series per year in English, although this figure might be considered very small by some informal learners who report watching between 300 and 400 hours of English language television per year, but these are probably exceptional cases. A corpus of 70 hours of viewing was therefore determined to be the most representative of the general experience of the students previously studied.

- 28 From the list of most frequently viewed series, five were chosen to represent a mixture of more serious and more humorous styles spanning the genres of soap opera and situation comedy. The series selected were a season of *House*, a season of *How I Met Your Mother*, a season of *Lost*, a season of *One Tree Hill* and a season of *Desperate Housewives*. A season generally consists of 22 to 24 episodes.
- 29 The way in which fans of certain series interact with them online is rarely studied in applied linguistics. Three such activities are fan fiction, in which fans write imaginary episodes of their favourite series, fan subtitling, in which amateur native language subtitles are provided for original version downloads, and fan transcription, in which fans of a series write up and put online a full transcript of episodes of their favourite series complete with stage directions, credits and other details. The existence of fan transcripts for the series chosen in this study (e.g., *twiztv.com* or *tvtdb.com*) greatly reduced the time taken to establish the corpus.
- 30 Care was taken to distinguish between parts of the script actually heard by viewers and stage directions and other items in the transcripts. All of these other items were removed from scripts in order to establish the actual quantity of dialogue in the various episodes. The transcripts also contained some typographical errors such as “dont” in place of “don’t” which required attention.
- 31 The HHOLD corpus was transformed into text format and analysed with Wordsmith 5. The cluster function was used to produce a ranking of the most frequently occurring clusters of four words. The 50 most frequently occurring clusters, with a minimum frequency of 17 are listed in Appendix 1. This situates the frequency per million words at 34. As in the Chen and Baker (2010) study, overlapping of clusters was taken into account so that the same groups of words were not counted twice. This process is particularly relevant in the case of the 5-gram “what are you doing here?” (a question routinely used to begin a scene in a television show) which is represented twice in the clusters and ranked first and seventh.
- 32 In order to establish the relevance of these clusters for language learning, they were analysed and characterised according to function as suggested in the Threshold Level (1976), and in the different levels of the CEFRL (2001).
- 33 The study further sought to establish the representativity of these frequently occurring clusters with respect to English as it is spoken in a broader context, as mirrored in the BNC. The frequency of these clusters in the spoken section of the BNC was therefore determined, and the clusters ranked by frequency in that corpus. The two rankings, that of the HHOLD corpus and the ranking of the same clusters in the BNC were then

compared using the Spearman rank correlation coefficient to determine whether these clusters appeared in comparable frequency order in a corpus more representative of English as it is spoken in a variety of contexts. In a final calculation, a Pearson correlation coefficient was used to determine whether the scaled frequency of use of these clusters in the BNC was similar to the HHOLD corpus. These calculations were carried out using online statistical tools available at <http://wessa.net/>. Since the spoken BNC corpus is some twenty times larger than the HHOLD corpus, values obtained from the BNC corpus were divided by twenty before comparison with HHOLD.

Results

- 34 A preliminary finding of this study is that the 70 hours of programming taken to represent the annual viewing of an average non-specialist student corresponds to some 100 different episodes, including a mixture of comedies (which run for 20 minutes) and dramas (which run for 42-44 minutes). The total number of words in the corpus is of the order of 500,000, and such quantitative elements are important when seeking to situate this research within the CREED framework which sees frequency of exposure to target structures as a significant factor in language learning.
- 35 The theoretical framework of the study would also point to the existence of recurrent structures in the corpus, and this is indeed the case. The 50 most frequently occurring 4-grams all appear a minimum of 17 times, suggesting that learners would be exposed to these lexical bundles more than once a month, with the top six being encountered on a weekly basis. These include items well known to English teachers as being problematic for French learners, such as “I want you to”, (and not the literal French translation: “I want that you”) and questions involving inversion and the use of auxiliaries where spoken French uses statements with rising intonation, hence “What do you think?” (and not “you think what?”). There are therefore some grounds for optimism regarding learning for some of these categories.
- 36 The predominant 4-word structures in the corpus (see Appendix 1) are examples of Threshold Level volitional language such as “I want you to”, “you want me to” and “I don’t want to.” This group is closely related to the CEFRL skills of expressing a wish or point of view, which can also be seen in structures such as “I thought you were,” “you don’t have to” or “was just trying to.”
- 37 These structures correspond well to the CEFRL B1 descriptor for spoken production: “I can connect phrases in a simple way in order to describe experiences and events, my dreams, hopes and ambitions. I can briefly give reasons and explanations for opinions and plans. I can narrate a story or film and describe my reactions.” (CEFRL 2001: 26)
- 38 Almost all of the 50 most frequently occurring structures are pronoun-verb structures, 24 of them using the verbs *to do*, *to want to*, *to have to*, *to need* and *to know*. These structures are of course action-related and coherent with the CEFRL view of the learner as a social agent: “it views users and learners of a language primarily as ‘social agents,’ i.e. members of society who have tasks (not exclusively language-related) to accomplish in a given set of circumstances, in a specific environment and within a particular field of action.” (CRFRL 2001: 9)
- 39 Almost half of the 50 most frequent structures appear in the context of questions such as “what are you doing?” or “what do you want?” which suggests that negotiation

frequently takes place in the series which make up the corpus. Indeed most dramatic situations can be seen as resembling information-gap tasks, in which information known to the viewer is gradually discovered by the characters. Hence these structures are easily transferable by learners to other communicative contexts. This negotiation is again reflected in the CEFRL B1 level descriptors, most clearly in the spoken interaction section: “I can deal with most situations likely to arise whilst travelling in an area where the language is spoken. I can enter unprepared into conversation on topics that are familiar, of personal interest or pertinent to everyday life.” (CEFRL 2001: 26)

- 40 A well-defined subset of the ranking list is made up of expressions containing swear words acceptable on prime-time American television, mostly involving the insertion of “the hell” into other structures, such as “what the hell is,” or “the hell are you,” or the single expression “son of a bitch,” which along with “no, no, no, no” is the only structure not to contain a verb. While it is important to identify these structures, their relative rarity as frequently-occurring structures is worth underlining. The presence of only five such structures in Appendix 1 should reassure teachers who may consider that there is little useful language to be learned from popular television series.
- 41 When comparing the frequency ranking of the 50 clusters in Appendix 1 with that of the same clusters in the BNC, a significant correlation can be observed. Figure 1 shows the correlation between the HHOLD corpus and the BNC and indicates a Spearman correlation coefficient of 0.314, which is significant at the 0.05 level for 50 pairs of rankings. Although not significant at the 0.01 level, this correlation suggests that the relative frequencies of use of these clusters are similar in the corpus series to that of real-life spoken English.
- 42 A number of points do differ greatly in ranking. This is illustrated by the two diagonal lines in Figure 1 which situate 12 of the 50 items outside the general correlation. Many of these outliers are the prime-time acceptable swear words mentioned above. It may however be interesting to analyse two terms from the list which differ greatly in relative frequency between the two corpora for reasons other than editorial regulation. The most over-represented term in the HHOLD list as indicated is “I need you to,” which is the least frequently-occurring of the 50 HHOLD clusters in the BNC, but is ranked 10th in the HHOLD list. Closer investigation of this cluster reveals that it is most frequently found in the medical drama *House*, as a standard manner of addressing patients. “I need you to roll up your sleeve” would be an example of this. It is therefore possible to see some influences of the professional substrate of this type of fiction in the HHOLD corpus.

Figure 1

- 43 The most under-represented term in the ranking list is “I don’t know how” which is ranked in the bottom 10 of the HHOLD list but in the top 10 of the corresponding BNC frequencies. It may be argued that the overwhelming presence of negotiation and volitional language in the HHOLD ranking list is an indication that while the scripts of these series are helpful illustrations of communication and negotiation of meaning, they more rarely present tasks being carried out and therefore make less reference to discussions of competency.
- 44 While the Spearman calculation compares ranks, it does not calculate the correlation between the number of occurrences.
- 45 In a second calculation, a Pearson correlation coefficient was calculated in order to determine the correlation between frequency of occurrence of the most frequent clusters in HHOLD and in the spoken section of the BNC. The correlation obtained is 0.206, which is not statistically significant. This outcome underlines the limited nature of the language used in the HHOLD corpus, and in particular the presence of the formulaic introductions to scenes with the words “what are you doing here?” and formulaic swearing structures as discussed earlier. Elimination of these structures from the calculation, as well as “I need you to,” for reasons discussed above, yields a correlation coefficient of 0.345 which is significant at the 0.02 level.
- 46 These results by no means prove that frequently occurring chunks of language found in American television series of recent years occurred with the same frequency in (largely British) spoken English in the late 20th century, as found in the BNC; indeed such a conclusion would be highly surprising. However they do show that these chunks are relevant target structures for threshold level learners and that, in most cases, their frequency is broadly similar to that found in real life conversations.

Limits and perspectives

- 47 This study of cluster frequency in a corpus of television series typically viewed by informal learners of English over twelve months constitutes a field of enquiry which is complementary to the main body of our research into the Online Informal Learning of English. The study seeks to characterise the input experienced by learners rather than

seeking to map input to intake. As such, further studies could be carried out into the degree of mastery of frequently occurring clusters from the corpus among informal learners of English. Most of the available corpora of OILE productions tend to be in written form, such as contributions to Facebook or productions relating to work in English classes, such as oral presentations. A forthcoming study (Sokkett, forthcoming) of blogs written by students of language didactics in which they analyse their own informal learning is likely to yield a great deal of information about how learners process language heard in television series and seek to re-use it. Testing groups of frequent and non-frequent viewers of the series studied in this research on their ability to use the most frequently occurring clusters would represent a further opportunity to gain insights into how learning is taking place.

- 48 The corpus itself is limited in size to a quantity corresponding to a typical year of viewing for the learners involved in this research. Given the availability of transcriptions of television series and other listening material used in OILE activities, a larger corpus could be developed and contrasts drawn between genres such as situation comedy and drama, or contexts such as medical or legal situations. The analysis in this study is limited to 4-grams, or 4-word clusters, according to the methodology of Chen and Baker (2010). Further studies of 3 and 5-word units may yield further insights into frequently occurring chunks of language and their role in informal learning. Finally, the study does not seek to characterise the way in which the structures are delivered such as accent, speed and background noise, which are aspects best analysed through learner self-report.

Conclusions

- 49 The aim of this study was to assess the relevance of watching original version television series for the informal learning of English. The overwhelming presence of frequently occurring everyday language at the B1 threshold level in the HHOLD corpus suggests that such viewing is indeed helpful for proceduralising relevant structures, and this assertion is supported by learner comments in our other studies (e.g., Sokkett & Toffoli, forthcoming) as discussed earlier. The extent to which these activities will contribute to an increase in listening comprehension levels in France should become visible over the next five to ten years as learners in their teens begin to spend more time watching television series in English. It is interesting to note that the development of terrestrial digital television has led to the first case of original version series being broadcast in France. The Disney Break programme on the NRJ12 channel is the first to broadcast teen series such as *Hannah Montana* in English without subtitles, no doubt as a reaction to the widespread downloading and streaming of such series by the target audience.
- 50 Teachers seeking to build on these new practices in their teaching are faced with a number of difficulties. They may be wary of creating a digital divide between learners who watch this type of output and those who do not. They may be concerned that downloading and streaming such files is illegal, although the teaching profession was deeply involved in (illegal) video cassette recording and copying in the 1980s and 1990s. Teachers may also feel at a disadvantage in terms of their multimedia skills, which are often less well developed than those of students. More generally, teachers may feel that popular fiction and comedy are not worthy representatives of the English language and

culture they were taught in university syllabuses which continue to focus on high culture.

- 51 An important first step towards taking account of informal listening in the classroom is to establish a dialogue with learners about their viewing and listening habits. Considering and treating learners as users of English can lead to a changed relationship between teachers and learners who are now not only able to offer examples of target structures from their listening activities (even though these may often be simple or approximate), but are also in search of strategies to help them improve their comprehension and enjoyment.
- 52 Allowing scope for learner participation in the choice of documents used in class is a simple way of recognising the status of language user which informal learning brings with it. Although some teachers may fear being undermined in their authority in the classroom by the use of documents which are mostly relevant to the learner, the potential for improvement in participation and focus on target forms may prove worthwhile in the long term.

BIBLIOGRAPHY

- Bonnet, Gérard (ed.). 2002. "The assessment of pupils' skills in English in eight European countries: A European project". Report downloadable from <<http://www.eva.dk/projekter/2002/evaluating-af-faget-engelsk-i-grundskolen/projektprodukter/assessmentofenglish.pdf>>.
- Chapelle, Carol. 2009. "The Relationship between second language acquisition theory and computer-assisted language learning". *The Modern Language Journal* 93, 741-753.
- Chen, Yu-Hua and Paul Baker. 2010. "Lexical bundles in L1 and L2 academic writing". *Language Learning and Technology* 14/2, 30-49. <<http://llt.msu.edu/vol14num2/chenbaker.pdf>>.
- Common European Framework of Reference for Languages* (CEFRL). 2001. Language Policy Division. Council of Europe, Strasbourg. Cambridge: Cambridge University Press.
- Cross, Jay. 2006. *Informal Learning*. San Francisco, CA: Pfeiffer.
- Ellis, Nick (ed.). 1994. *Implicit and Explicit Learning of Languages*. London: Academic Press.
- Ellis, Nick. 2002. "Frequency effects in language processing". *Studies in Second Language Acquisition* 24, 143-188.
- Ellis, Nick. 2007. "The associative cognitive creed". In VanPatten, Bill & Jessica Williams (eds.), *Theories in Second Language Acquisition*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Gramsci, Antonio. 1991. *Selections from Cultural Writings*. Cambridge, MA: Harvard University Press.
- Krashen, Stephen D. 1982. *Principles and Practice in Second Language Acquisition*. Oxford: Pergamon.
- Larsen-Freeman, Diane and Lynne Cameron. 2008. *Complex Systems and Applied Linguistics*. Cambridge: Cambridge University Press.

Legallois, Dominique and Jacques François. 2006. "Autour des grammaires de constructions et de patterns". *Les Cahiers du CRISCO* 21. Université de Caen. <www.crisco.unicaen.fr/IMG/pdf/cahier21.pdf>.

Rieder, Angelika. 2003. "Implicit and explicit learning in incidental vocabulary acquisition". *IEWS* 12-2. University of Vienna. <http://www.univie.ac.at/Anglistik/views/03_2/RIE_SGLE.PDF>.

Rupérez Micola A., A. Bris and A. Banal-Estañol. 2009. TV or Not TV? Subtitles and English Skills. Available at SSRN <<http://ssrn.com/abstract=1403964>>.

Sinclair, John. 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

Sockett, Geoffrey. 2011. "Les processus cognitifs de résolution de problèmes pour l'apprentissage des langues dans des environnements multimedia : Apprentissage informel et réseaux sociaux". *Les Cahiers de l'Acedle* 8/1, 29-46.

Sockett, Geoffrey. Forthcoming. "La complexité au service de l'apprentissage informel des langues. Communication au Colloque EPAL 2011 <http://w3.u-grenoble3.fr/epal/dossier/03_pro/resumes_epal.pdf>.

Sockett, Geoffrey and Denyze Toffoli. Forthcoming. "Beyond learner autonomy: A dynamic systems view of the informal learning of English in virtual online communities". *Proceedings of EuroCALL 2010*. <http://eurocall.webs.upv.es/index.php?m=menu_02&d=actas>.

Stevens, Anne. 2010. "Étude sur l'impact des technologies de l'information et de la communication (TIC) et des nouveaux médias sur l'apprentissage des langues". Commission Européenne <http://eacea.ec.europa.eu/llp/studies/documents/study_impact_ict_new_media_language_learning/final_report_fr.pdf>.

Swain, Merrill. 2005. "The output hypothesis, theory and research". In Hinkel, E. (ed.), *Handbook of Research in Second Language Teaching and Learning*. Mahwah, NJ: Lawrence Erlbaum, 471-482.

The Threshold Level for Modern Language Learning in Schools. 1976. London: Council of Europe/Longman.

Tissot, Philippe. 2004. "Terminology of vocational training policy. A multilingual glossary for an enlarged Europe". Luxembourg: CEDEFOP.

Toffoli, Denyze and Geoffrey Sockett. 2010. "How non-specialist students of English practice informal learning using web 2.0 tools". *ASp* 58, 125-144.

Van Geert, Paul. 2008. "An introduction to complex dynamic systems". *Modern Language Journal* 92/2, 179-199.

Vanderplank, Robert. 2010. "Déjà vu? A decade of research on language laboratories, television and video in language learning". *Language Teaching* 43/1, 1-37.

Wright, Beverly, Andrew T. Sabin, Yuxuan Zhang, Nicole Marrone and Matthew B. Fitzgerald. 2010. "Enhancing perceptual learning by combining practice with periods of additional sensory stimulation". *Journal of Neuroscience* 30/38, 12868-12877.

Websites used in this study (All sites consulted on 16th March 2011.)

Davies, Mark. 2004-. BYU-BNC: British National Corpus. <<http://corpus.byu.edu/bnc/>>.

Deezer. <<http://www.deezer.com>>.

La Coccinelle. <<http://www.lacoccinelle.net>>.

Le blog Musique. http://www.leblogmusique.net/tops_tlchargements/index.html.

Scott, Mike. 2008. WordSmith Tools version 5. Liverpool: Lexical Analysis Software <www.lexically.net>.

Spotify. <<http://www.spotify.com>>.

Twist TV. <<http://www.twitv.com>>.

TV transcript data base. <<http://www.tvtdb.com>>.

Wessa, P. 2011. Free Statistics Software, Office for Research Development and Education, version 1.1.23-r7 <<http://wessa.net>>.

APPENDIXES

Appendix 1
Ranking and frequency of the 50 most frequently occurring 4-word clusters in the HHOLD corpus

Rank	Cluster	Freq.	Rank	Cluster	Freq.
1	WHAT ARE YOU DOING	158	26	NOTHING TO DO WITH	24
2	I WANT YOU TO	65	27	SO WHAT DO YOU	24
3	YOU WANT ME TO	63	28	YOU DON'T WANT TO	24
4	WHAT DO YOU THINK	48	29	YOU KNOW WHAT I	23
5	I DON'T WANT TO	47	30	DO YOU WANT ME	22
6	WHAT DO YOU WANT	47	31	I AM SO SORRY	22
7	ARE YOU DOING HERE	41	32	IF YOU WANT TO	22
8	NO NO NO NO	41	33	WHAT DID YOU DO	22
9	YOU DON'T HAVE TO	41	34	DO I HAVE TO	21
10	I NEED YOU TO	40	35	I WAS TRYING TO	21
11	WHAT THE HELL IS	38	36	YOU HAVE NO IDEA	21
12	THE HELL ARE YOU	36	37	DO YOU HAVE A	20
13	I DON'T KNOW WHAT	34	38	I NEED TO KNOW	20
14	HOW DO YOU KNOW	30	39	CAN I HELP YOU	19
15	WHAT DO YOU MEAN	30	40	COME ON COME ON	19
16	ARE YOU GOING TO	28	41	DO YOU WANT TO	19
17	I JUST WANTED TO	28	41	HELL ARE YOU DOING	19
18	SON OF A BITCH	28	43	GIVE IT TO ME	18
19	WHAT THE HELL ARE	28	44	CAN I GET YOU	17
20	I THOUGHT YOU WERE	27	45	GOING TO HAVE TO	17
21	WHERE ARE YOU GOING	27	46	I DON'T KNOW HOW	17
22	I HAVE TO GO	25	47	I NEED YOUR HELP	17
23	ARE YOU ALL RIGHT	24	48	KNOW WHAT TO SAY	17
24	ARE YOU TALKING ABOUT	24	49	WAS JUST TRYING TO	17
25	GET OUT OF HERE	24	50	WHAT ARE YOU TALKING	17

NOTES

1. <<http://www.academie-francaise.fr/langue/questions.html#anglicismes>>
2. Salience concerns how readily noticeable an item is to the learner. Examples of salience might include a plural form which changes the number of syllables in a word.
3. Data of this kind is available online, at sites such as <<http://www.leblogmusique.net/>>.

ABSTRACTS

Research into the online informal learning of English often seeks to describe what target-language related activities learners engage in in their spare time and what effect these might have on formal learning. The cultural hegemony of English in the mass media in France means that these activities are commonplace amongst students majoring in subjects other than English. Informal listening activities include downloading or streaming English language television series, however the relevance of the language content of these activities remains to be proven. In this study, a corpus of television series scripts corresponding to a year of viewing by non-specialist students is constructed and analysed using a concordancer. The most frequently occurring 4-word clusters are collated and characterised according to their linguistic content. These clusters, ranked by frequency, are compared to the frequency of the same clusters in the British National Corpus. A correlation is established between these two rankings, which suggests that the most frequently occurring clusters in the present corpus are indeed relevant examples of target language.

La recherche sur l'apprentissage informel de l'anglais en ligne porte souvent sur la description des loisirs d'apprenants ayant un lien avec l'anglais et leur impact sur l'apprentissage formel. L'hégémonie culturelle de l'anglais dans les médias de masse en France fait que ces pratiques sont fréquentes parmi des étudiants spécialistes de disciplines autres que l'anglais. Le téléchargement ou visionnage en streaming de séries télévisées en version originale figure souvent parmi les activités informelles d'écoute, mais la pertinence des contenus langagiers de ces émissions pour l'apprentissage de la langue-cible reste à démontrer. Dans cette étude, nous construisons un corpus de transcriptions de séries télévisées correspondant à une année de visionnage par un étudiant du secteur LANSAD. Le corpus est analysé afin de faire ressortir les groupes de quatre mots les plus fréquents. Nous analysons le contenu linguistique de ces constructions et le classement par fréquence est ensuite comparé avec celui des mêmes constructions dans le British National Corpus. La corrélation entre ces deux classements permet de conclure que les constructions les plus fréquentes du corpus constituent des exemples pertinents de la langue cible.

INDEX

Keywords: associative-cognitive approach, cluster frequency, CREED, informal learning, television series

Mots-clés: apprentissage informel, approche associationniste-cognitive, fréquence de collocations, série télévisée

AUTHOR

GEOFFREY SOCKETT

Geoffrey Sockett is a member of the language didactics team of the LILPA research unit (EA 1339) at the University of Strasbourg and director of a Master's degree in multimedia design for language learning and teaching. He teaches in the fields of psycholinguistics, task-based approaches to language learning and online language learning. His research interests include the above areas and relate particularly to informal learning. gsockett@unistra.fr