

Recurrent Multiword Sequences in L2 English Spoken Academic Discourse: Developmental Perspectives on 1st and 3rd Year Undergraduate Presentational Speech

Nicole Baumgarten, University of Southern Denmark

Abstract

Little is known about the development of second language (L2) capacities in L2 users located in multilingual environments where more than one language is a viable communication tool and users can decide which to use for which purpose. Adopting a socially-grounded perspective on L2 learning, this study explores university-based L2 English learning in a multilingual learning context in Denmark. 46 academic presentations from first and final year undergraduate students were analysed for the users' use of recurrent multiword sequences as a measure of development of routinized discourse production. This analysis was complemented by an analysis of the users' language use habits and socialization patterns. The analyses revealed a negative development in pre-patterned L2 use between first and final year students. These results call for a reconsideration of academic L2 English instruction in multilingual environments outside native English-speaking settings, where L2 learning trajectories appear to be able to be stunted by L2 users' overall language use habits.

Keywords: Multiword sequences, lexical bundles, English as a lingua franca, L2 socialization, L2 learning, university language, monologic speaking

1. Introduction

Socially-grounded theories of second language (L2) learning have promoted conceptual shifts from a focus on the L2 *learner* to the L2 learner as an L2 *user* and from an exclusive focus on the L2 as solely located in an individual's mind towards the systematic recognition of the situatedness of the L2 learning and usage process in very specific social environments which interconnect in various ways with the affective and cognitive aspects of L2 learning (e.g. Kramsch, 2000; Firth and Wagner, 1997, 2007). In this way, L2 learning becomes a factor of the socio-cultural context in which it occurs, i.e. the setting, people, purposes, topics, registers, genres and other languages involved as well as the interactional norms that regulate the communicative encounters L2 users find themselves in. The context places affordances on each individual user, and each user responds to these affordances in different ways which

Baumgarten, Nicole. 2014. "Recurrent Multiword Sequences in L2 English Spoken Academic Discourse: Developmental Perspectives on 1st and 3rd Year Undergraduate Presentational Speech." *Nordic Journal of English Studies* 13(3):1-32.

2 *Nicole Baumgarten*

eventually become evident in the L2 learning outcome. Individual responses to contextual affordances can be explained by recourse to individual personality factors in L2 learning or by recourse to the users' subjective understanding of their L2 learning and usage environment and their place within it. The latter approach lays focus on the environmental exigencies as perceived by the users and their relevance for users' agency in the process of L2 learning, shaping their investment in learning and the processes of identity formation and self-actualization bound up with it.

Taking as a point of departure the notion that, from the perspective of an L2 user, L2 learning is a socially-embedded event, this study investigates L2 development in spoken academic English through an apparent time cross-sectional study of authentic spoken discourse by adult advanced L2 users in university settings. The L2 users investigated are enrolled in an undergraduate trilingual (Danish, English, German) Humanities programme at a university in Denmark. This university is characterized by a high number of international staff and students and international study programmes. The university campus is located in the Danish-German border region and draws a sizable number of students from Germany. Substantial student numbers also come from Eastern Europe and Asia. Next to the local language Danish, English and German are regularly used by students and staff; a variety of other languages are used for non-official purposes among ethnic student groups. In this space, languages are in constant competition regarding their usefulness for individual speakers. Against this background, we assume that the uses and the usefulness of a particular language in this particular environment influences individuals' language use habits and L2 learning outcomes. Focusing on the students' L2 academic English, the purpose of the investigation is to verify whether development in L2 use can be meaningfully linked to individual users' subjective perceptions of what their L2 learning context is like and to what extent information about this context can serve as an explanation for L2 development.

For the purpose of the present analysis, naturally occurring first year and final year language production data were used to compare entry and exit level L2 English performance in one specific register: the academic oral presentation. The analysis of the language data focuses on the use of recurrent multiword sequences as an indicator of L2 development in

terms of pre-patterned language use and routinization of discourse production. The production data are contextualized by self-report data collected from the participants through survey questionnaires and interviews, providing insight into each individual's subjective learning context regarding their idea of their own L2 achievement level, membership in individual social networks and day-to-day language use patterns.

The use of recurrent multiword sequences has increasingly come to be considered as an important measure of L2 development in adults' spoken English (Adolphs and Durow, 2004; Crossley and Salisbury, 2011; De Cock, 2000; Wood, 2006; Ellis et al., 2008; Qi and Ding, 2011; Sánchez-Hernández, 2013). Most of the available studies have, in their research design, delimited their perspective in specific ways, favouring elicited or experimental data (De Cock, 2000; Ellis et al., 2008), focusing on L2 development in native English-speaking settings (Adolphs and Durow, 2004; Crossley and Salisbury, 2011; Wood, 2006), or employing limited social data to contextualize the observed language use patterns (an exception is Adolphs and Durow, 2004). A comprehensive inclusion of contextual constraints on L2 learning is often beyond the scope of these investigations as is the analysis of L2 use in naturalistic settings. The present study attempts to combine the analysis of language use and learning context data in order to shed light on L2 learning outcomes as a socially embedded phenomenon. The results of this investigation can feed into the planning and evaluation of educational policy and practice for academic learner populations in internationalized educational settings outside native English contexts.

The following section situates this investigation within language socialization theory, summarizes the special characteristics of L2 socialization into English and outlines the concept of L2 development used. Section 3 defines the notion of recurrent multiword sequences adopted for this investigation and explains their relation to L2 development. Section 4 describes the language and self-report data used and the analytical procedures applied. Sections 5 and 6 present the major results of the analysis and discuss them in relation to the major facets of the participants' learning contexts as gleaned from the self-report data. Section 7 concludes the article by considering some implications of the investigation for university-level L2 English instruction.

2. Background

2.1 Language socialization

The perspective on L2 learning adopted here is grounded in language socialization theory (Ochs, 1986), where language development is seen as culturally situated, reverberating with factors beyond language, e.g. politics, economics and attitudes. Learning occurs in expert-novice interactions which provide two interrelated types of knowledge and learning: (1) Learners acquire the language through direct and indirect teaching of what to say and write in specific contexts (“socialization to use language”, Ochs, 1986, pp.2-3). (2) Learners are engaged in a process of “socialization through the use of language” (ibid.). This refers to the acquisition of the socio-cultural meaning of linguistic forms in a given situation and the way they pattern the social interaction, the statuses of the participants and their role relationships in community-specific ways.

L2 socialization proceeds similarly along direct and indirect teaching in interpersonal constellations and the simultaneous enculturation into L2 community-specific social practices. It differs from L1 socialization in at least four respects (Dufon, 2008), however, which render the socialization trajectory comparatively unpredictable. First, L2 learners have restricted access to target language native speakers for input and relevant target culture interaction. Second, the socialization process results in multilingual persons who are members of more than one speech community. Their community affiliations may be only partial, resulting in less than full adoption of community-specific interactional practice and native-like acquisition and control of the relevant linguistic and socio-pragmatic knowledge.¹ Third, when L2 socialization is predominantly classroom-based, input and interaction are restricted to the situations and roles available in the classroom setting. Further, the national and cultural systems that characterize the learning institution will shine through in classroom practice and consequently be part of the L2 socialization. Fourth, L2 socialization into *English* might be substantially different from socialization into other L2s because English is currently caught up in both localizing and internationalizing processes (cf. Crystal, 2010) which lead to a considerable variability of forms and

¹ This process will also occur in multilingual L1 socialization.

usage conventions in the language (cf. Seidlhofer, 2004; Canagarajah, 2007).

When English L2 socialization takes place outside a native English speaking context, the socialization process takes on additional distinctive features: L2 English may be approached by learners as an exceptionally useful lingua franca, enabling international as well as social, educational and professional mobility. In this role, English may have only limited potential as a language for individual identification and affiliation with a larger community construct against which L2 socialization takes place (House, 2003; Pavlenko and Norton, 2007; Higgins, 2011). A second unique feature relates to the experts that socialize the novice user. These experts might be themselves L2 speakers of various L1 backgrounds, and their expert status might be coupled with various levels of L2 knowledge (Dufon, 2008). Crucially, the relationship between language use and situational appropriateness might have to be characterized in fundamentally different ways in L2 English socialization processes in which predominantly L2 users are involved. Research on the use of English as a vehicular language (i.e. lingua franca use) has shown, e.g., that infelicitous language use is normalized in interaction and that interactional trouble based on linguistic choice occurs only rarely (House, 1999; Firth, 1996, 2009). A shared sense of situational appropriateness that translates into a normative sense of linguistic choice in the realization of interactional patterns does not seem to exist and mechanisms regulating linguistic and interactional behaviour seem to be largely suspended. It is, therefore, possible that the community against which L2 socialization into English takes place is perceived by learners as unconnected to a clearly definable and geographically locatable group of expert language users which shares an extensive linguistic repertoire and uses this in predictable ways for the realization of a relatively stable set of social practices. This instability of the notion of L2 English community opens up a multitude of possible L2 socialization trajectories (Duff, 2007) and leads to an ‘individualistic appropriation’ of the L2 which finds expression in L2 use (Kramsch, 1998).

2.2 L2 development

In multilingual environments, L2 development is susceptible to both endemic notions of linguistic and communicative norms and L2 users’

needs-driven choices in their L2 learning and usage environments. Individuals' conceptions of linguistic norms are less consistent in multilingual settings than in monolingual settings (Le Page and Tabouret-Keller, 1985). Notions of appropriate choice as well as shared knowledge of the sociopragmatic meaning potential of linguistic expressions may be less likely to be taken for granted across language users. In particular for the use of L2 English as a lingua franca, it has been argued that whenever there is little or no exposure to native register norms these are also of little real normative value for L2 users (Canagarajah, 2007; Seidlhofer, 2004; Kalocsai, 2008), which complexifies for these users the understanding of the relationship between linguistic expertise, functional identity choices in the community and attendant L2 learning needs.

For these reasons, English L2 development in multilingual environments outside a native English context must be framed in special ways. In these environments, a focus on L2 development in the direction of target language (TL) norms and 'learner language' development towards native speaker-like capacities is not useful: TL norms and native-like proficiency might not be interactionally relevant in their environment and infelicitous language use might not be procedurally consequential. In order to describe L2 development under the conditions of learning and use in a multilingual setting, it needs to be understood broadly as change that occurs in the linguistic forms employed in a given communicative situation. In the present context, L2 development is conceptualized as *change* in the frequency of use of linguistic forms. A quantitative difference in the linguistic choices users make is taken to reflect a difference in situation-specific L2 performance as a result of learning.

3. Recurrent multiword sequences

3.1 Definition

This study investigates recurrent multiword sequences (RMS)² as markers of L2 development. RMS have been investigated under a variety

² The term RMS is used throughout as a cover term referring to the general phenomenon of recurrent word strings. Where necessary, the concepts used by individual authors have been retained.

of names, including lexical phrases (Nattinger and DeCarrico, 1992), routine formulae (Coulmas, 1979), conversational routines (Aijmer, 1996), recurrent word combinations (Altenberg, 1998), lexical stems (Pawley and Syder, 1983), lexical bundles (Biber et al., 2004), formulas (Simpson-Vlach and Ellis, 2010), and formulaic sequences (Wray, 2002). Research can be roughly separated into two main strands: *formulaic language* and *lexical bundles*. Research on formulaic language usually assumes some measure of collocational strength between the words that make up a sequence; the lexical bundles approach is a purely frequency-based approach to RMS.

This investigation adopts the lexical bundles approach to RMS because the use of statistical measures of collocation (e.g. *Mutual Information*) in the identification of RMS in language use biases the analysis towards sequences of lexical words by disregarding high frequency function words (Biber, 2009). In addition, collocational strength only reflects the likelihood of co-occurrence of lexical words within a span of words but not the likelihood of a particular sequence of words (Biber, 2009). In contrast, a frequency approach identifies highly frequent sequences of both lexical and function words and is, thus, better able to reflect actual language use patterns.³

Under the lexical bundles approach, RMS are defined as the most frequently occurring sequences of words in a register (Biber et al., 2004; Biber and Barbieri, 2007; Biber, 2009). Lexical bundles are usually not idiomatic in meaning, not perceptually salient, and often do not constitute a complete structural unit. They can be continuous or discontinuous, are usually transparent in meaning, and have identifiable discourse functions in the register in which they occur. Because they are frequency phenomena, the lexical bundles of a register can only be discovered through corpus-driven⁴ analyses. High frequency is taken to

³ Cf., however, Groom (2009) for an analysis of L2 writing that combines a lexical bundles approach with measures of collocational strength; Granger and Paquot (2008) propose suggestions for reconciling different methodological traditions in phraseological research.

⁴ The term is used following Biber (2009). It is defined as an inductive approach to the analysis of language corpora proceeding from only minimal assumptions about (the validity of) linguistic constructs. Linguistic constructs emerge from the analysis. Thus, grammatical classes and syntactic structures have no a-priori status in corpus-driven analyses.

reflect the pre-patterned or formulaic status of a bundle (Biber and Barbieri, 2007): The bundle is stored in the lexicon as an unanalysed multiword chunk and, because of its non-compositionality, does not present production difficulties for the speaker (Biber et al., 2004).

3.2 RMS as indicators of L2 development

RMS occur in users' language output as a function of time and exposure to a register. In L2 learning contexts, RMS are typically investigated as markers of lexical and sociopragmatic development or register-specific competence, usually in comparison to an adult native speaker norm (Pawley and Syder, 1983; Wray and Perkins, 2000; 3.3 below). RMS are also taken to play a central role in the development of L2 fluency (Wood, 2006).

RMS can also be considered from the perspective of routinization of L2 discourse production. On this view, RMS are a measure of language development because RMS are evidence of "routinized ways of presenting information in continuous discourse" (Altenberg, 1998: p. 121; also Biber et al., 2004). RMS develop in users' output because speakers deal with recurrent communicative tasks linguistically in the same way so that a pattern of linguistic choice becomes established. RMS have further been described as the building blocks of continuous discourse. Altenberg (1998) and Biber et al. (2004) show that continuous discourse consists of interlocking sequences of RMS and that they function as structural frames in discourse followed by an open slot for the presentation of new information (e.g. *I don't know if* + new information). In this way, RMS are scaffolding devices for discourse production. The presence of RMS in language use, thus, indicates degrees of routinization of discourse production. Differences in the frequency of RMS across RMS types, speakers or time indicate differences in the degree of routinization and the expression of the discourse functions associated with the RMS.

The investigation of RMS from the point of view of routinization of discourse production additionally allows one to assess the L2 use of RMS without reference to a native speaker norm. The occurrence of RMS themselves, native-like or not, reflects routinization of language use and signals the presence of building blocks for the construction of continuous discourse.

3.3 L2 use of RMS

Most of the research on adult L2 RMS use is concerned with comparisons of L1 and L2 use. The analyses show that L2 users overuse, underuse and ‘misuse’ RMS (e.g. De Cock, 2000; Ellis, et al. 2008; Durant and Schmitt, 2009; Sánchez-Hernández, 2013). L2 users’ sets of RMS include non-native-like sequences (partly influenced by their L1s) and overlap only partially with the set of L1 users. One problem for L2 users seems to lie in the difficulty of establishing what constitutes a viable L2 RMS (Wray and Perkins, 2000; Wood, 2006). Ellis et al. (2008) find that, in the identification of RMS, L2 users seem to rely more on frequency information than on the collocational strength between the words in the sequences. In their experiments on academic formula use, L2 users showed better knowledge of high than low frequency RMS. Ellis et al. (2008) explain this as a result of the fact that L2 users encounter fewer and only high frequency RMS, which makes them consider frequency information when they assess the pre-patterned status of a sequence of words. This is in contrast to L1 users, who encounter all RMS of their L1 frequently, which leads them to consider “underlying co-occurrence information” (p. 301) revealing the association between the words that make up the sequence. These findings support the assumption that frequency of use reflects that a given RMS has status as a viable pattern for L2 users.

There is evidence that L2 users’ repertoire of RMS broadens with length of L2 exposure (Adolphs and Durow, 2004; Wood, 2006, Bardovi-Harlig and Bastos, 2011; Crossley and Salisbury, 2011; Qi and Ding, 2011; Sánchez-Hernández, 2013). The increase in RMS, however, does not necessarily coincide with an increase in proficiency in terms of grammatical correctness or an increase in vocabulary (Yorio, 1989; Qi and Ding, 2011). It also does not necessarily coincide with an increase in only native-like RMS (Wray, 2002). This suggests that, for L2 users, routinization of discourse production is accomplished more easily than lexicogrammatical progress and development towards target-like usage norms, which, in turn, underpins the idea that development of RMS use outside a consistent norm-providing target language environment may follow special paths. The fact that L2 RMS increase over time, however, shows that the effects of RMS use, namely routinized production of continuous discourse, is one central learning outcome.

Wray (2002) argues that L2 acquisition of formulaic sequences is dependent on the L2 user's circumstantial and personal needs. For example, users' L1 can hinder L2 formula acquisition when the L1 is perceived as a legitimate and useful means of communication in the L2 setting. Likewise, L2 users' identity choices can undercut their desire to achieve a high degree of assimilation into a L2 community, reflected in the retention of non-native-like formulaic sequences. As a consequence, L2 users' sets of RMS are likely to be individualistic, reflecting their experience in L2 learning and use.

For the present analysis, the acquisition and use of RMS is conceived of as a needs-based process. RMS constitute a repertoire of pre-patterned language choices which speakers deem functional in the context of use and which reflect the degree of routinization of language use in a given situation. The analysis focuses on the use of RMS in academic presentational monologue as one register of academic speaking. These extended speaking turns (Coulthard and Montgomery, 1981) have been identified as high stakes academic communicative situations for L2 speakers (Graham and Barone, 2001; Mauranen, 2009) because they require confident, effective and intelligible speech production over a prolonged period of time under real-time production constraints. As such, academic presentations are instances in which pre-patterned language and routinization of discourse production help speakers to manage the affordances of presentation situation (Graham and Barone, 2001; Nesi and Basturkmen, 2009).

4. Data and method

This is a naturalistic enquiry into L2 learning, seeking a community-specific understanding of L2 RMS usage patterns. The data comprise classroom-based L2 performance data as an objective measure of L2 use and development and L2 users' self-report data as subjective assessments of their L2 learning and usage environment. The L2 users investigated are students in a trilingual 3-year undergraduate program in Modern Languages and Intercultural Communication at a small-size university in Denmark. They are predominantly L1 speakers of Danish or German; a small number (<5) are German-Danish bilinguals. In this programme, students enrol in the first year with a minimum of high-intermediate to low-advanced L2 English (equivalent to IELTS 6.5) and advanced

beginner/low intermediate levels for L2 Danish and German. English proficiency and English-medium classes run throughout the programme, requiring students to give several academic presentations in English each semester.

The language data consist of two sets of academic student presentations from the first and third (final) years of the study program. These presentations are obligatory but ungraded components of a language proficiency module *English spoken and written language use*. The learning objective of this module is the acquisition of advanced academic discourse and business communication skills. The module consists of 4 consecutive 13-week courses in semesters 1, 2, 3, and 5 of the programme. The presentations used for this investigation come from one first and one fifth semester course. Both courses were conducted by the same instructor. The students in the fifth semester course had had all previous courses of the module with that instructor as well. In total, 46 presentations were video-recorded in fall 2011 and afterwards manually transcribed using an orthography-based transcription model (HIAT, Rehbein et al., 2004). The sizes of the resultant YEAR1 and YEAR3 corpora are given in Table 1.

Table 1. Corpus information

	YEAR1	YEAR3
Number of presentations/ speakers	27	19
Presentation topic	Free choice among topics of 'personal interest'	Free choice among 'current controversial issues'
Number of words	35.257	10.676

The *clusters/ngrams*-function of the concordancer software AntConc (Anthony, 2012) was used to extract all RMS from the two corpora. The cut-off frequency was set at three occurrences⁵, i.e. to be counted as an RMS each sequence had to recur at least three times in any one presentation or across presentations. This procedure identified RMS shared by more than one speaker in each group as well as idiosyncratic

⁵ Large-scale corpus studies of lexical bundles have used higher frequency limits to identify register-specific use (e.g. Biber et al., 2004). In small corpora, representing the language use of groups of limited size, a high cut-off captures only very few, highly indexical RMS, which may not be an accurate reflection of actual RMS usage in the group or any one speaker.

language use. From the total of RMS, those occurrences were eliminated which were specific to the topic of a presentation and recurred in that presentation only (e.g. “the Rønsdam waterwork”). The final sets of RMS were then analysed for their frequency distribution within and across the groups, their structural patterns, and discourse functions.

The self-report data were collected in parallel with the recording of the presentations and targeted the L2 users’ socializing patterns in personal social networks, language usage habits and self-assessments of their L2 achievement level at the time of their academic presentation. Participation in this data collection was voluntary. The data were elicited in individual structured interviews conducted by student peers and from survey questionnaires administered by the researchers during class sessions. Only those students for whom both language and self-report data were available were included in the analysis.

5. Results

5.1 RMS

5.1.1 General frequencies

Figures 1 and 2 present the normed frequency counts⁶ of RMS types and tokens in YEAR1 and YEAR3 presentations. As a point of numerical comparison to academic presentations by English L1 users, the figures include the frequency of RMS in the Arts and Humanities lectures component of the British BASE⁷ corpus and monologic discourse in the American MICASE⁸ corpus.

⁶ Throughout normed frequencies per 1,000 words are used.

⁷ The British Academic Spoken English corpus project, Universities of Warwick and Reading.

⁸ Michigan Corpus of Academic Spoken English, University of Michigan.

Figure 1. RMS types

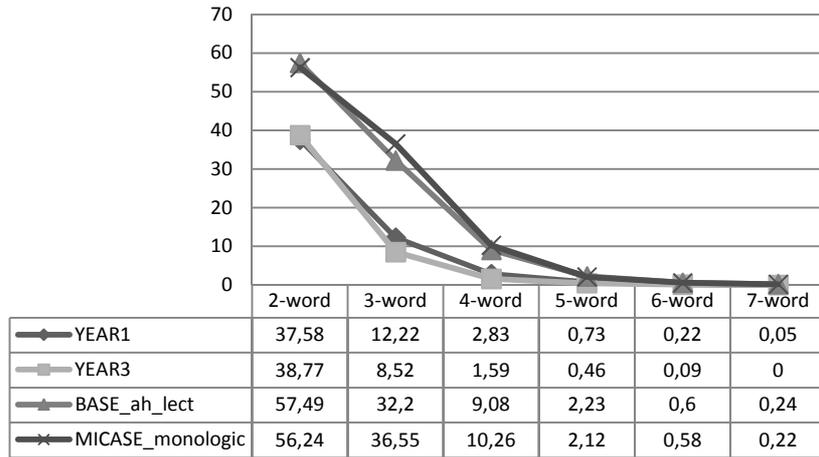
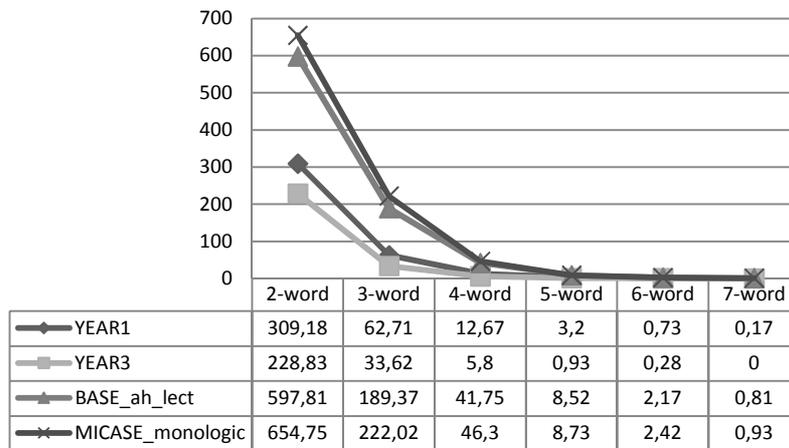


Figure 2. RMS tokens



The frequency of RMS types is similar across YEAR1 and YEAR3. While there is a minimally higher frequency of 2-word RMS in YEAR3, the frequencies for all longer RMS spans are somewhat lower than in YEAR1. Overall, the L2 users have considerably fewer RMS across all RMS spans than L1 users.

Across all spans, RMS token use is higher in YEAR1 than in YEAR3. The overall token use in the L2 data is considerably lower than in the L1 data. L2 users control a comparatively limited set which does not appear to grow and whose use in talk seems to decrease with longer L2 exposure and more practice and experience in the register.

Figure 3 breaks down the data at the individual level, showing the percentages of the presentation talk realized by RMS. The L2 users in both years use RMS to different degrees, from between 5-10% of their talk to between 25-35%. RMS use plays a greater role in YEAR1 speaker's talk.

Figure 3. Percentage of RMS in individuals' talk

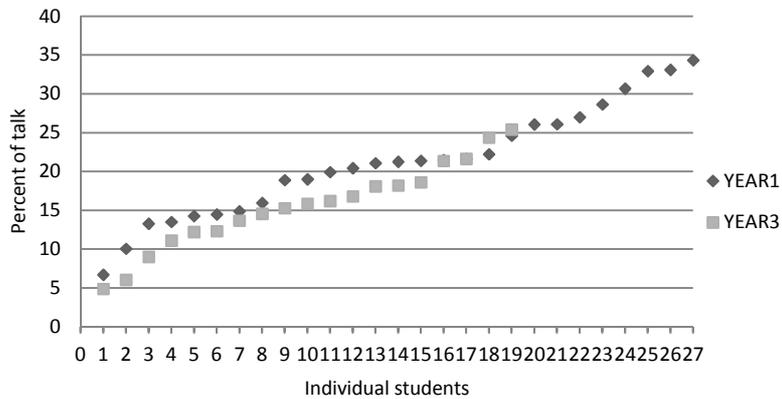
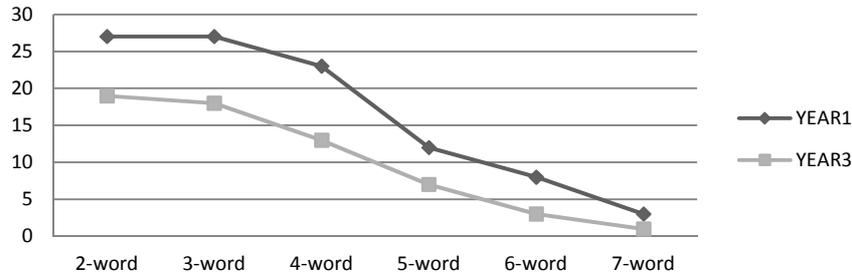


Figure 4 shows that the majority of speakers in both YEAR1 and YEAR3 produce 2-, 3- and 4-word RMS. Longer RMS are individual speaker phenomena in both groups. In the following, we will first consider the very frequent 2-word RMS before turning to the longer pre-patterned sequences.

Figure 4. Distribution of RMS sizes within groups



5.2 2-word RMS

There is a considerable frequency gap between 2-word and longer RMS, with 2-word RMS making up roughly 80 percent of the total of tokens in both groups. Table 2 shows the top 25 RMS for YEAR1 and YEAR3.

In YEAR1 all of the top 25 are shared by the majority of speakers; in YEAR3 16 are shared by a majority. 2-word RMS are the smallest units of pre-patterned language use. They are basic building blocks of discourse in that they facilitate clause and utterance initiation, rudimentary clause complexing, basic discourse connectivity as well as speaker and audience positioning. Five main structural patterns can be discerned:

1. PP-fragments; used in adverbials and prepositional objects (e.g. *And made its way to Europe **in the** twentieth century*).
2. Pronoun/existential *there+BE*; used as utterance and clause beginnings (e.g. ***This is** kind of an experience?*)
3. Fixed RMS with fixed metadiscourse functions, e.g. epistemic stance markers (*I think, kind of, of course*), vague quantifier *a lot*, elaboration maker *for example*
4. Coordinator *and+determiner/pronoun/connector*, used for phrasal and clausal coordination and discourse connectivity on utterance level (***And the** last things would be the faces of intoxication **and the** effects that occur when using crystal METH?*)

5. RMS with *you*; used to connect with the audience (e.g. *And if you compare it to to our situation it's I think unbelievable?*)

Table 2. Top 25 2-word RMS; shared RMS in bold face

YEAR1			YEAR3		
Frequency	No. of speakers	RMS	Frequency	No. of speakers	RMS
5.02	25	in the	5.34	18	of the
4.59	25	of the	3.75	15	in the
4.31	23	you can	3.65	11	i think
2.69	25	and the	3.09	12	on the
2.41	25	is a	2.81	15	and the
2.41	21	on the	2.72	12	have to
2.35	24	a lot	2.62	13	that the
2.18	22	to the	2.25	12	is that
2.18	17	this is	2.15	14	to be
2.16	21	is the	2.15	12	a lot
2.16	17	if you	2.15	8	would be
2.10	25	there is	2.06	12	it is
2.10	21	can see	2.06	12	to the
1.93	21	and they	2.06	7	of course
1.90	21	they are	1.87	6	the the
1.90	19	and then	1.78	14	there are
1.73	20	you have	1.69	6	for example
1.73	16	it is	1.59	11	lot of
1.65	20	there are	1.59	10	from the
1.62	21	it was	1.59	10	there is
1.59	21	i think	1.59	9	they are
1.47	16	which is	1.50	11	for the
1.45	19	lot of	1.41	11	with the
1.33	14	and you	1.41	9	think that
1.25	14	kind of	1.41	7	if you

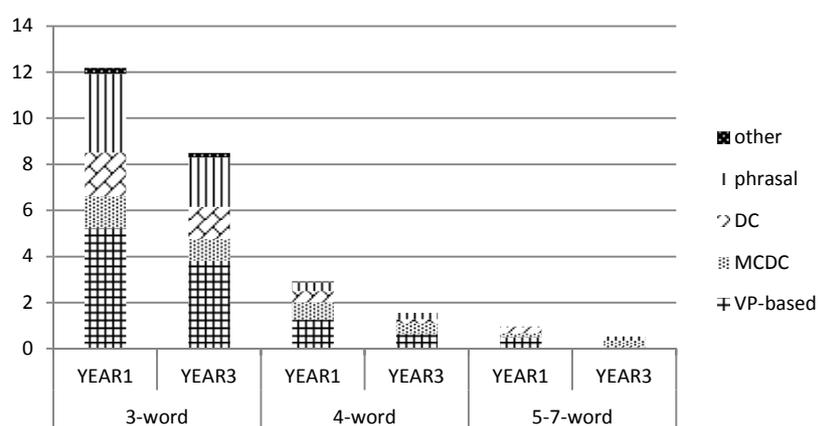
Among the 2-word RMS, YEAR1 has more clausal fragments, i.e. VP-based RMS, which facilitate automatized clause and utterance beginnings, whereas YEAR3 has a greater portion of prepositional RMS and RMS which bridge clause and phrase boundaries facilitating clause complexing (cf. Biber, 2009) (e.g. *There are documents that states/ that state that the earliest known description of a person that suffered from multiple sclerosis dates back to the fourteenth century in Holland; The second argument is that larger schools are much cheaper to to run or to maintain.*).

5.3 Longer RMS

5.3.1 Structure

The 3-7-word RMS⁹ were categorized according to their structural type following the typology proposed by Biber et al. (2004). The categories are VP-based (*you can see on*), main clause with dependent clause fragment (MCDC) (*I don't know if you*), dependent clause fragment (DC) (*which is the*), and phrasal fragment (*most of the*). Because of their limited occurrence of less than one per 1,000 words in both data sets, 5-7-word RMS were collapsed into one set. While the overall frequency of RMS is lower in YEAR3, YEAR1 and YEAR3 still show similar distributions of structural types for all RMS spans, as shown in Figure 5.

Figure 5. Structural patterns of RMS types



5.3.2 Discourse functions

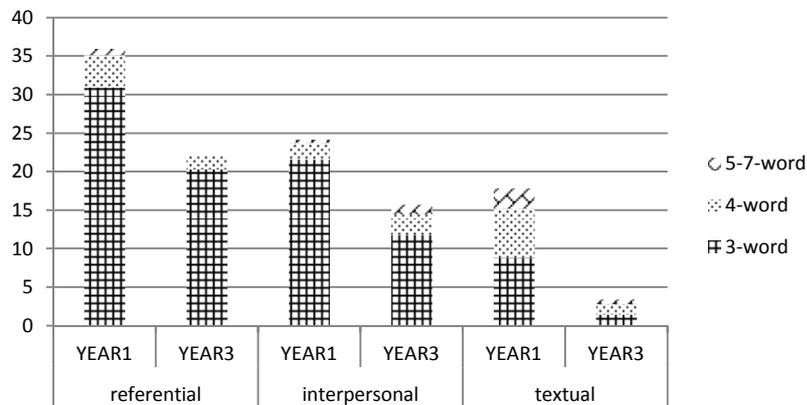
The longer RMS tokens were also categorized according to the discourse functions they fulfil in their contexts of occurrence. This categorization distinguished between *referential*, *interpersonal* and *textual* functions, elaborated from the categories suggested by Biber et al. (2004) and Simpson-Vlach and Ellis (2010). Referential RMS refer to concrete or abstract entities in the text in order to identify the entity or to specify

⁹ A table illustrating the 25 most frequent longer tokens is given in the Appendix.

some particular attribute of the entity (e.g. *You have the the mountains **in the middle of the country***). Interpersonal RMS express epistemic stance, attitudinal and modal meanings (e.g. *But it was amazing how much he learned in this **I don't know** about seven month we did it.*) as well interaction with the audience (e.g. *Any of **you have any questions?***). Textual RMS serves discourse structuring purposes (e.g. *So I'm not **going to talk about it now.***) and make reference to prior or upcoming discourse or entities in the extralinguistic context of the presentation situation (e.g. *That's what **you can see on the right/ in the right picture.***) This three-way categorization covers the three main tasks of presentational speaking, i.e. content presentation and topic specification, expression of speaker subjectivity and audience involvement, and multimodal discourse management.

Referential RMS are used most frequently in both groups, followed by interpersonal and textual RMS (Figure 6). YEAR3 use textual RMS comparatively infrequently. Especially for referential and textual RMS, YEAR1 have longer–4-7-word–RMS. YEAR3 have longer RMS in particular for the realization of interpersonal functions.

Figure 6. Discourse functions of RMS tokens*



*Unclear categorization occurred in a few cases of 3-word RMS, amounting to 1.27 and 0.28 per 1000 words for YEAR1 and YEAR3, respectively. The unclear cases are not included in the Figure.

In summary, development in RMS use cannot be described for the level of RMS *types*. With regard to RMS *tokens*, YEAR3 speakers use more of

the shortest (2-word) RMS than YEAR1, although the majority of YEAR3 speakers do have longer (3- and 4-word) RMS in their repertoire. There is no difference between YEAR1 and YEAR3 in the structural patterns preferred for RMS. In both groups, there is a general preference for VP-based RMS, which is a preference that has also been observed for L1 conversational English (Biber, 2009). The RMS fulfil slightly different functions in YEAR1 and YEAR3: In both groups RMS are used to facilitate content presentation and the expression of speaker stance and audience involvement. YEAR3 users make comparatively little use of RMS with discourse management functions, which leads one to speculate that either discourse organization is less frequently signalled in general or discourse management is achieved through singular, non-recurring expressions.

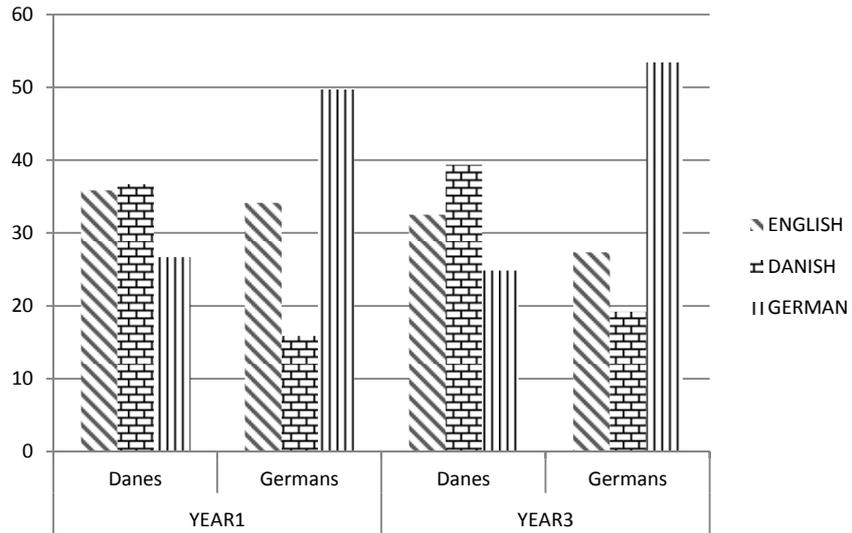
5.4 Language usage patterns

5.4.1 Distribution of languages in use

Figures 7 and 8 present the students' language use habits inside and outside university on an average day. The figures were calculated from L2 users' responses to the interview questions "If you think of a normal day during the semester, how are German, Danish and English distributed over the course of the day? Could you give percentages?" and "How is the distribution of languages in your spare time?"¹⁰

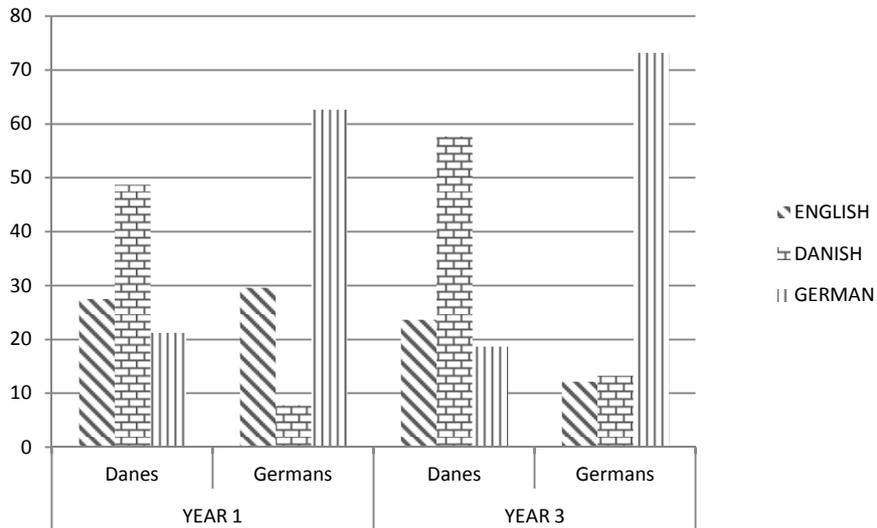
¹⁰ The use of additional languages occurred only infrequently and is not included in the Figures.

Figure 7. Language use inside university in percent*



*Labelling into *Danes* and *Germans* according to L1; bilinguals chose their label.

Figure 8. Language use outside university in percent



Language use inside and outside university is characterized by somewhat different patterns for Danish and German students. Compared to the Danes, the Germans in YEAR1 and YEAR3 show a clearer dominance of L1 use. For Danes and Germans, the use of L2 English is less frequent in YEAR3 than in YEAR1. Possibly, this is a direct consequence of the increase in L1 use, as personal networks become established and network-specific language choice determines language use. Social networks might become established in the first place because they allow specific language choice (see 5.4.2). In YEAR1, the role of English as lingua franca across many different communicative constellations, involving interactions with the international student body, the scientific and administrative staff, as well as with the local community outside the university seems to be responsible for the greater role of English for the German students in YEAR1. In general, English is used in L2-L2 constellations. Infrequent interaction with L1 speakers of English is reported by less than five students overall.

5.4.2 Social networks

Figures 9 and 10 represent the students' socializing patterns inside university. The social networks were drawn from the interview questions "Can you name the three people you talk most with inside uni?" and "Which languages do you use with each person". Information about the reciprocity and multiplexity of the network relations were elicited through further prompts. In the networks, double arrows represent reciprocal relationships, i.e. the relationship was named by both individuals. Single, double and triple arrows represent uniplex and two levels of multiplex ties. Uniplex ties represent a relationship on one dimension only, e.g. 'fellow student'; multiplex ties represent a relationship involving more than one social connection, e.g. 'fellow student and car-pooling'. Solid circles represent individuals from the same year of the program, dotted circles individuals from outside that group. 'G' and 'DK' identify German and Danish students, respectively. For each relationship, the preferred languages are given.

Figure 9. Social network YEAR1

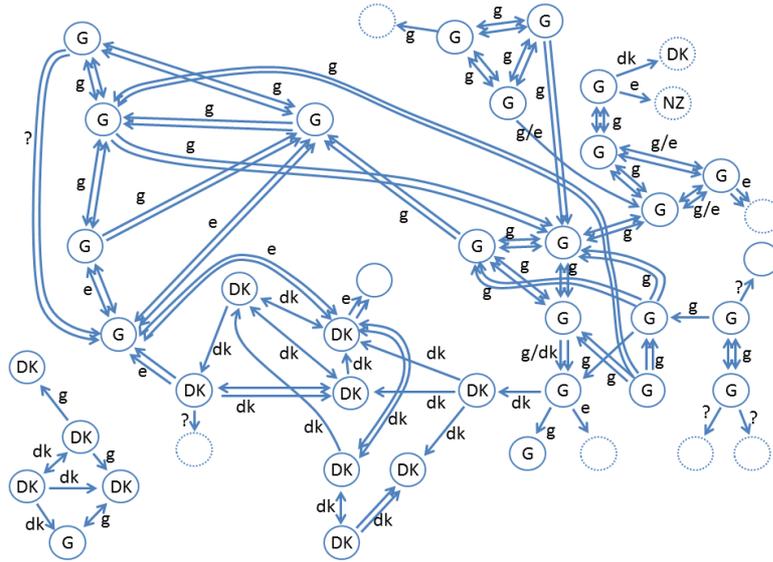
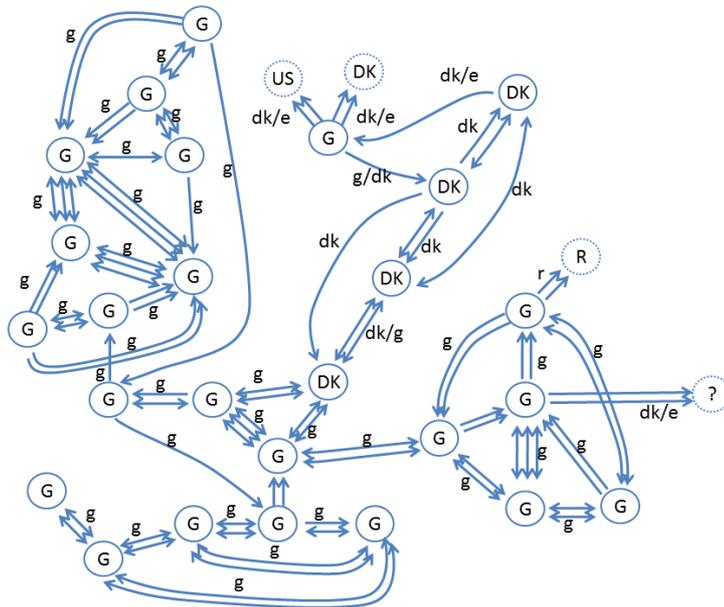


Figure 10. Social network YEAR3



The social networks show clearly defined groupings of people with L1-based language use for YEAR1 and YEAR3. Most relationships are with other students from the same year; most contacts are within one's L1 group. The interactions are predominantly carried out in the students' L1. In YEAR1, fewer relationships are reciprocal and the relationships are less complex. At the time of data collection, the students had known each other for about 2 months. The networks reflect this state of initial acquaintance. Still, there are closer network ties among the German students. They seem to be moving closer together than their Danish counterparts because, in contrast to these, the German students share halls of residence, form car-pools and, thus, have opportunity to socialize with each other outside university. The Danish students tend to live in established social circles of family and long-term friends, which to a lesser degree requires them to form new sustaining relationships upon starting their university studies. In YEAR3, the relationships tend to be multiplex among students from the same year, reflecting 2,5 years of highly frequent (minimum of three days for a minimum of four hours) contact.

5.4.3 L2 self-assessment

The self-assessment represents students' subjective rating of their English L2 capacities at the time of their presentations. Using a Likert-type scale format, the students were asked to rate their L2 abilities in academic and everyday language use in all four skills on a numerical scale between "not at all" (0) and "like a native speaker" (10). The self-ratings place YEAR1 and YEAR3 in the upper third of capacities. The ratings differ only minimally between YEAR1 and YEAR3, but with the exception of "reading newspapers/magazines" YEAR3's self-ratings are slightly lower than YEAR1's. The non-development in RMS use across YEAR1 and YEAR3, thus, seems to correspond to a non-development in the assessment of own L2 capacities. In both groups, general proficiency skills are rated higher than academic proficiency skills. Classroom-based speaking, of which presentational speaking is one part, receives the lowest ratings in both groups.

Table 3. L2 English self-assessment across skills

	YEAR1	YEAR3	Average deviation YEAR1	Average deviation YEAR3
SPEAKING – with friends	7.8	7.5	1.3	1.1
	– in class	7.3	6.9	1.3
WRITING – text chat. e-mail. internet	8.2	7.9	1.1	1.1
	– for class	7.6	6.9	1.1
READING – newspapers/magazines	8.2	8.4	0.9	0.8
	– scientific texts	7.3	7.2	1.2
LISTENING – lectures	8.2	7.9	1.1	1.2
	– TV series. reality shows	8.3	8.0	1.1

6. Discussion

The analyses of the L2 performance and learning context data show only few differences between YEAR1 and YEAR3. After three years of exposure to English, YEAR3 users do not show evidence of more pre-patterned language use in academic speaking than YEAR1 users. On the contrary, in YEAR3, the number of RMS types is lower, the use of the shortest RMS tokens is higher and RMS are used for a more restricted set of discourse functions, disfavours discourse organization. Taking pre-patterning of language as a measure of the degree of routinization in the production of continuous discourse in a situation, the use of fewer and shorter RMS in the nominally more advanced group has to be taken as a sign of no or reversed development. With only short pre-patterned sequences routinely available, ad-hoc compositional utterance planning is necessary almost permanently, making it comparatively effortful for speakers to hold the floor for extended periods of time. What is the reason for this non-evolved use of RMS in the final year of studies, despite L2 exposure and experience in the register?

In both groups, users claim for themselves L1-based socialization patterns within their group and a decrease in the use of English in their daily life. At the same time, their self-rating of L2 proficiency reveals their self-identification as advanced L2 English users, in particular, in regard to general non-academic language skills. This apparent discrepancy of L2 performance and language use habits in social practice, on the one hand, and self-assessment, on the other, might be explained by considering who models L2 English use for these users, i.e.

who leads their L2 socialization process. The social networks show that almost all daily interactions are among the students of the same year of the programme, which makes them rather close-knit groups. It is also an artefact of the campus at which they study that each year moves through the program as a fixed group of students. Throughout the three years there is no fluctuation in the group. As most interactions at the university are carried out in the students' L1s, the situations in which academic English is used regularly are restricted to the classrooms. Consequently, on the occasions where academic English has to be used in monologic talk the users exclusively encounter each other as models. In the absence of modelling by expert users or opportunities for comparison with users from outside the group, 'advanced' self-ratings on the basis of exposure to only peer-to-peer 'novice' language use appear reasonable.

Wray (2000) states that the 'poverty of the L2 experience' often prevents the acquisition of RMS because RMS are either not present in the L2 input or they are present in receptive contexts only (e.g. non-interactive media) where the lack of meaningful interaction impedes RMS acquisition. Schmidt (1990) puts forth that 'noticing' of linguistic structures is necessary before a structure is successfully taken in and acquired, and Ellis et al. (2008) point out that L2 users predominantly notice high frequency items in input and consequently acquire mainly these. When the students predominantly have each other as models for L2 English use in presentational talk, only the shorter RMS will be frequent enough to be noticeable. In other words, the limited use of RMS in the input is likely to influence, over the years, a development of RMS use biased towards the short, highly frequent elements, i.e. those which occur frequently in every presentation. This limited set, then represents the common ground of situationally appropriate pre-patterned language use. As the relationships between the students in the group develop greater complexity, monologic performance in the L2 might also become less risky for a user's identity in the group, because it might have become less susceptible to L2 performance factors, but sustained by features of the user's relationships with others in the group, who invariably make up the audience of the L2 use. Regarded from a needs-driven perspective on L2 learning, the reduced set of RMS and the minimal degree of automatization of discourse production and basic discourse management that come with it are, then, adequate to sustaining a self-identity in the group. Although it is possible that factors beyond this analysis are

involved in the explanation of RMS use in the present data, L2 usage habits, socialization patterns and attitudes towards own L2 capacities are part of the combination of factors which lead to what could be called arrested development for YEAR3 L2 users.

7. Conclusion

Although this investigation is cross-sectional and does not predict academic L2 socialization trajectories in academic L2 learning, all groups of students go through the same sequence of courses, which requires them to perform in the academic oral presentation genre at regular intervals. From the point of view of the educational institution, then, the absence of a clear difference in L2 capacities in terms of the routinization of L2 performance between students in year one and students in year three is a problematic diagnosis, although the absence of an increase in routinization does not necessarily imply an absence of increase in general proficiency. It does, however, imply that in internationalized educational settings where English plays a substantial—but not exclusive—role, L2 users do not acquire a greater degree of routinization of L2 use simply as a factor of time and exposure. The results of this investigation, thus, raise the question of the role and capacity of English L2 instruction in multilingual university settings outside a native English speaking context.

In a context that is characterized by almost exclusive L2 use of English and in which L1 use is a legitimate option—such as in places which for geographic or economic reasons attract particular ethnonational student groups—English L2 instruction acquires a new central role. This role must be based on a re-alignment of learning objectives with the conditions of the learning context. First, instruction needs to balance out L2 users' other, non-academic L2 English experiences as well as their multilingual repertoire. Second, it needs to counteract the 'poverty of the L2 experience' which presents a quantitative disadvantage in exposure to pre-patterned language use. Finally, L2 classroom instruction, and all activities connected to it, need to be consistently conceived of as the main agent in the academic L2 socialization process because no other models might be available, putting L2 users at a risk of underdevelopment of their academic communication skills.

Acknowledgements

This research is in part funded by the BHJ foundation, Denmark (Svend Beck, Peter Holm og Vagn Jacobsens Almene Fond).

References

- Adolphs, Svenja and Valerie Durow. 2004. "Social-cultural integration and the development of formulaic sequences." *Formulaic sequences: Acquisition, processing and use*. Ed. Norbert Schmitt. Amsterdam: John Benjamins. 107-126.
- Aijmer, Karin. 1996. *Conversational routines in English: Convention and creativity*. London/New York: Longman.
- Altenberg, Bengt. 1998. "On the phraseology of spoken English: The evidence of recurrent word combinations." *Phraseology: Theory, analysis and applications*. Ed. A. P. Cowie. Oxford: Oxford University Press. 101-122.
- Anthony, Laurence. 2012. AntConc.[AntConc 3.3.5w]. Available from <http://www.antlab.sci.waseda.ac.jp/software.html>
- Bardovi-Harlig, Kathleen and Maria Thereza Bastos. (2011). "Proficiency, length of stay, and intensity of interaction and the acquisition of conventional expressions in L2 pragmatics." *Intercultural Pragmatics* 8/3: 347-384.
- Biber, Douglas and Francesca Barbieri. 2007. "Lexical bundles in university spoken and written registers." *English for Specific Purposes* 26: 263-286.
- Biber, Douglas. 2009. "A corpus-driven approach to formulaic language in English. Multi-word patterns in speech and writing." *International Journal of Corpus Linguistics* 14/3: 275-311.
- Biber, Douglas, Susan Conrad and Viviana Cortes. 2004. "If you look at ...: Lexical bundles in university teaching and textbooks." *Applied Linguistics* 25/3: 371-405.
- Canagarajah, Suresh. 2007. "Lingua franca English, multilingual communities, and language acquisition." *The Modern Language Journal* 91/1: 923-939.
- Coulmas, Florian. 1979. "On the sociolinguistic relevance of routine formulae." *Journal of Pragmatics* 3/3-4: 239-266.

- Coulthard, Malcolm and Martin Montgomery. 1981. "The structure of monologue." *Studies in Discourse Analysis*. Ed. Malcolm Coulthard and Martin Montgomery. London: Routledge and Kegan Paul. 31-38.
- Crossley, Scott and Thomas Lee Salisbury. 2011. "The development of lexical bundle accuracy and production in English second language speakers." *International Review of Applied Linguistics* 49: 1-26.
- Crystal, David. 2010. "The future of Englishes: Going local." *From international to local English – and back again*. Eds. Roberta Facchinetti, David Crystal and Barbara Seidlhofer. Frankfurt am Main: Peter Lang. 17-26.
- De Cock, Sylvie. 2000. "Repetitive phrasal chunkiness and advanced EFL speech and writing". *Corpus Linguistics and Linguistic Theory*. Papers from the Twentieth International Conference on English Language Research on Computerized Corpora (ICAME 20) Freiburg im Breisgau 1999. Eds. Christian Mair and Marianne Hundt. Amsterdam: Rodopi. 51-68.
- Duff, Patricia A. 2007. "Second language socialization as sociocultural theory: Insights and issues." *Language Teaching* 40: 309-319.
- Dufon, Margaret A. 2008. "Language socialization theory and the acquisition of pragmatics in the foreign language classroom." *Investigating Pragmatics in Foreign Language Learning, Teaching and Testing*. Eds. Eva Alcón Soler and Alicia Martínez Flor. Bristol: Multilingual Matters. 25-44.
- Durant, Philip and Norbert Schmitt. 2009. "To what extent do native and non-native writers make use of collocations?" *International Review of Applied Linguistics* 47: 157-177.
- Ellis, Nick C., Rita Simpson-Vlach and Carson Maynard. 2008. "Formulaic language in native and second language speakers: Psycholinguistics, corpus linguistics, and TESOL." *TESOL Quarterly* 42/3: 375-396
- Firth, Alan. 1996. "The discursive accomplishment of normality: On 'lingua franca' English and conversation analysis". *Journal of Pragmatics* 26/2: 237-259.
- Firth, Alan. 2009. "The lingua franca factor." *Intercultural Pragmatics* 6/2: 147-170.
- Firth, Alan and Johannes Wagner. 1997. "On discourse, communication, and (some) fundamental concepts in SLA research." *The Modern Language Journal* 81/3: 285-300.

- Firth, Alan and Johannes Wagner. 2007. "Second/foreign language learning as a social accomplishment: Elaborations on a reconceptualized SLA." *The Modern Language Journal* 91/1: 800-819.
- Graham, Janet G. and Susan M. Barone. 2001. "Academic speaking: Learning to take 'longer turns'." *Understanding the courses we teach: Local perspectives on English language teaching*. Eds. John Murphy and Patricia Byrd. Ann Arbor: University of Michigan Press. 281-308.
- Granger, Sylviane and Magali Paquot. 2008. "Disentangling the phraseological web." *Phraseology. An interdisciplinary perspective*. Eds. Sylviane Granger and Fanny Meunier. Amsterdam: John Benjamins. 27-50.
- Groom, Nicholas. 2009. "Effects of second language immersion on second language collocational development." *Researching collocations in another language – Multiple interpretations*. Eds. Andy Barfield and Henrik Gyllstad. Basingstoke: Palgrave Macmillan. 21-33.
- Higgins, Christina. 2011. "The formation of L2 selves in a globalizing world." *Identity formation in globalizing contexts: Language learning in the new millennium*. Ed. Christina Higgins. Berlin: Walter de Gruyter. 1-18.
- House, Juliane. 1999. "Misunderstanding in intercultural communication. Interactions in English as lingua franca and the myth of mutual intelligibility." *Teaching and learning English as a global language*. Ed. Claus Gnutzmann. Tübingen: Stauffenberg Verlag. 73-92.
- House, Juliane. 2003. "English as a lingua franca: A threat to multilingualism?" *Journal of Sociolinguistics* 7/4: 556-578.
- Kalocsai, Karolina. 2008. "Erasmus exchange students: A behind-the-scenes view into an ELF community of practice." *Apples – Journal of Applied Language Studies* 3/1: 25-49.
- Kramsch, Claire. 1998. *Language and culture*. Oxford: Oxford University Press.
- Kramsch, Claire. 2000. "Second language acquisition, applied linguistics, and the teaching of foreign languages." *The Modern Language Journal*. 84/3: 311-326.

- Le Page, Robert B. and Andrée Tabouret-Keller. 1985. *Acts of identity: Creole-based approaches to language and ethnicity*. Cambridge: Cambridge University Press.
- Mauranen, Anna. 2009. "Spoken rhetoric. How do natives and non-natives fare?" *Cross-linguistic and cross-cultural perspectives on academic discourse*. Eds. Eija Suomela-Salmi and Fred Dervin. Amsterdam: John Benjamins. 199-218.
- Nattinger, James F. and Jeanette DeCarrico. 1992. *Lexical Phrases and Language Teaching*. Oxford: Oxford University Press.
- Nesi, Hilary and Helen Basturkmen. 2009. "Lexical bundles and discourse signalling in academic lectures." *Lexical cohesion and corpus linguistics*. Eds. John Flowerdew and Michaela Mahlberg. Amsterdam: John Benjamins. 23-44.
- Ochs, Elinor. 1986. "Introduction." *Language socialization across cultures*. Eds. Bambi B. Schieffelin and Elinor Ochs. Cambridge: University of Cambridge Press. 1-13.
- Pavlenko, Aneta. Bonnie Norton. 2007. "Imagined Communities, Identity, and English Language Learning." *International Handbook of English Language Teaching*. Eds. Jim Cummings and Chris Davison. New York: Springer. 669-680.
- Pawley, Andrew and Frances Hodgetts Syder. 1983. "Two puzzles for linguistic theory: Nativelike selection and nativelike fluency." *Language & Communication*. Eds. Jack C. Richards and Richard W. Schmidt. London/New York: Longman. 191-226.
- Qi, Yan and Yanren Ding. 2011. "Use of formulaic sequences in monologues of Chinese EFL learners." *System* 39: 164-174.
- Rehbein, Jochen, Thomas Schmidt, Bernd Meyer, Franziska Watzke and Annette Herkenrath. 2004. „Handbuch für das computergestützte Transkribieren nach HIAT“. *Arbeiten zur Mehrsprachigkeit, Folge B* 56: 1-90.
- Sánchez-Hernández, Purificación. 2013. "Lexical bundles in three oral corpora of university students." *Nordic Journal of English Studies* 12/1, 187-209.
- Schmidt, Richard W. 1990. "The Role of Consciousness in Second Language Learning." *Applied Linguistics* 11/2: 129-158.
- Seidlhofer, Barbara. 2004. "Research perspectives on teaching English as a lingua franca." *Annual Review of Applied Linguistics* 24: 209-239.

- Simpson-Vlach, Rita and Nick C. Ellis. 2010. "An academic formulas list: New methods in phraseology research." *Applied Linguistics* 31/4: 487-512.
- Wood, David. 2001. "In search of fluency. What is it and how can we teach it?" *The Canadian Modern Language Review / La revue canadienne des langues vivantes* 57/4: 573-589.
- Wood, David. 2006. "Uses and functions of formulaic sequences in second-language speech: An exploration of the foundations of fluency." *The Canadian Modern Language Review / La revue canadienne des langues vivantes* 63/1: 13-33.
- Wray, Alison and Michael R. Perkins. 2000. "The functions of formulaic language: An integrated model." *Language & Communication* 20: 1-28.
- Wray, Alison. 2000. "Formulaic sequences in second language teaching: Principle and practice." *Applied Linguistics* 21/4: 463-489.
- Wray, Alison. 2002. *Formulaic language and the lexicon*. Cambridge: Cambridge University Press.
- Yorio, Carlos. A. 1980. "Conventionalized language forms and the development of communicative competence." *TESOL Quarterly* 14/4: 433-442.
- Yorio, Carlos A. 1989. "Idiomaticity as an indicator of second language proficiency." *Bilingualism across the lifespan: Aspects of acquisition, maturity and loss*. Eds. Kenneth Hyltenstam & Loraine K. Obler. Cambridge: Cambridge University Press. 55-72.

Appendix

Top 25 3-7word RMS; shared RMS in bold face

YEAR1			YEAR3		
Frequency	No. of speakers	RMS	Frequency	No. of speakers	RMS
2.01	20	you can see	1.59	11	a lot of
1.42	19	a lot of	0.94	8	i think that
0.68	13	i don't know	0.75	6	on the other
0.68	9	i'm going to	0.75	5	the other hand
0.62	15	there is a	0.66	5	on the other hand
0.60	13	i want to	0.56	6	but i think
0.57	10	one of the	0.56	5	argument is that
0.51	9	going to talk	0.56	5	you have to
0.48	9	as you can	0.56	4	i think it's
0.48	9	this is the	0.56	2	to say that
0.45	9	in the world.	0.47	4	they have to
0.45	8	going to talk about	0.47	4	would be that
0.43	8	don't know if	0.47	3	in order to
0.43	9	as you can see	0.47	3	one of the
0.40	11	to talk about	0.47	3	there will be
0.40	10	can see the	0.47	2	is of course
0.40	9	i think it's	0.47	2	it would be
0.40	8	and you can	0.47	2	of the opinion
0.37	9	and this is	0.47	2	the fact that
0.37	4	if you go	0.47	2	the opinion that
0.37	9	you can see the	0.47	2	they want to
0.37	7	i don't know if	0.47	1	have to say
0.37	7	i'm going to talk	0.47	2	of the opinion that
0.34	9	you have to	0.47	1	have to say that
0.34	7	can see it	0.37	4	but on the
0.34	7	some of the	0.37	4	i think it
0.34	6	i'm going to talk about			
0.34	7	some of the			