Univerzita Karlova v Praze

Filozofická fakulta

Ústav anglického jazyka a didaktiky

Filologie – Didaktika konkrétního jazyka (anglický jazyk)

Tomáš Gráf

Accuracy and fluency in the speech of the advanced learner of English

Přesnost a plynulost v mluveném projevu pokročilých studentů angličtiny

Disertační práce

Vedoucí práce – prof. PhDr. Aleš Klégr

2015

Poděkování

– manželce Sáře a synovi Jakubovi za to, že stáli při mně;

– svému školiteli, Aleši Klégrovi, za veškerou podporu, rady a povzbuzení;

– Janu Čermákovi za nekonečnou inspiraci;

– Ondřeji Tichému a Markétě Malé za to, že se včas ocitli na správném místě a navrhli mě jako koordinátora českého korpusu LINDSEI;

– mateřskému Ústavu anglického jazyka a didaktiky za pomoc, podporu, kolegialitu a přátelství;

– studentům ÚAJD za to, že se nechali nahrát a pomohli s přepisy;

– Tereze Štěpánové, Veronice Pojarové a Lucii Jiránkové za obětavou pomoc se sestříháním nahrávek LOCNEC;

– Ústavu Českého národního korpusu za finanční podporu projektu LINDSEI_CZ;

– Fonetickému ústavu FF UK za zapůjčení nahrávacího studia;

– Centru pro anglickou korpusovou lingvistiku Katolické univerzity v Lovani a Katedře anglistiky Univerzity Justuse Liebiga v Giessenu, jmenovitě Gaëtanelle Gilquinové a Sandře Götzové, za pohostinnost a odbornou pomoc při studijních pobytech a konzultacích;

– Sandře Götzové za inspiraci svou knihou o plynulosti;

– Sylviane Grangerové za přelaskavé svolení použít Lovaňskou rukověť chybového značkování;

– všem přátelům, které jsem nudil rozpravami o probíhajícím výzkumu;

– tátovi, že mi celá léta naznačoval, jakým směrem a k jakému cíli jít, a věřil, že tam jednou dojdu.

Prohlašuji, že jsem disertační práci napsal samostatně s využitím pouze uvedených a řádně citovaných pramenů a literatury a že práce nebyla využita v rámci jiného vysokoškolského studia či k získání jiného nebo stejného titulu.

Tomáš Gráf

V Praze dne 24.6.2015

Abstrakt

Disertační práce analyzuje přesnost a plynulost mluveného projevu pokročilých mluvčích angličtiny, studentů anglistiky. Data čerpá ze žákovského korpusu sestávajícího z padesáti patnáctiminutových rozhovorů s těmito studenty a z paralelního korpusu obsahujícího čtyřicet devět stejně dlouhých rozhovorů s rodilými mluvčími angličtiny. Přesnost v projevu českých mluvčích je zkoumána technikou chybové analýzy. Tak jsou identifikovány hlavní rysy pokročilé žákovské angličtiny, přičemž následné kvantitativní analýzy odhalují, že v celé skupině studentů, jejíž jazyková úroveň je nečekaně široká, se dva základní typy chyb vyskytují častěji než chyby jiné. Jsou to chyby v užití členů a slovesných časů. K popisu plynulosti je vybrán vzorek proměnných – tempo mluvy a frekvence vyplněných i nevyplněných pauz – a výsledky jsou porovnány s paralelním korpusem rodilých mluvčích. Ti mluví rychleji než většina ze zkoumaných studentů. Studenti nadužívají vyplněné i nevyplněné pauzy a produkují kratší úseky řeči mezi pauzami. Korelace mezi přesností a plynulostí nebyla na vzorku prokázána. Disertace je první analýzou takto rozsáhlého vzorku českých pokročilých mluvčích angličtiny. V závěru jsou navržena četná pedagogická a metodologická východiska.

Klíčová slova: přesnost, plynulost, chybová analýza, tempo mluvy, chybovost, pauzologie, jazyková pokročilost, mluvený jazyk, produkce jazyka

Abstract

The thesis analyses the accuracy and fluency exhibited in the spoken advanced-learner English of Czech students of English philology. It draws its data from a learner corpus comprising fifty 15-minute interviews with these learners and from a parallel native-speaker corpus of forty-nine 15-minute interviews. As regards accuracy, the learner data is analysed using techniques of error analysis. Salient features of advanced learner English are identified and the subsequent quantitative analyses reveal that throughout the entire group of students (which is characterized by what revealed itself to be a wide proficiency span) two groups of error types are found to be much more frequent than any other, namely errors in the use of articles and tenses. For the fluency measurements a small selection of variables has been chosen to describe speed fluency (speech rate) and breakdown fluency (the frequency of unfilled and filled pauses), and the results are compared with those for the parallel native-speaker corpus. The analysed native speakers are found to produce speech at a generally much higher rate than the majority of the learners. There does not appear — at least in the light of the given sample — to be any direct correlation between fluency and the frequency of errors. Moreover, the learners are found to overuse filled and unfilled pauses and to produce shorter speech runs. The study provides the first analysis of such a large sample of Czech advanced learners of English. The conclusion of the thesis offers numerous pedagogical and research implications.

Keywords: accuracy, fluency, error analysis, speech rate, error rate, pausology, advanced language proficiency, spoken language, language production

Contents

List of tables

List of figures

List of abbreviations*

CAF — complexity, accuracy and fluency
CEA — computer-aided error analysis
CIA — contrastive interlanguage analysis
CLT — communicative language teaching
CZ followed by a number (e.g. CZ001) – numeric speaker identification in LINDSEI_CZ
EN followed by a number (e.g. EN001) – numeric speaker identification in LOCNEC
ER — error rate
FP — filled pause
GTM — grammar-translation method
ICLE — International Corpus of Learner English
L1 — first language, mother tongue
L2A — second language acquisition (the process of)
L2 — second/third language learnt
LINDSEI — Louvain International Database of Spoken English Interlanguage
LOCNEC — Louvain Corpus of Native English Conversation
MLR — mean length of runs
NS — native speaker
NNS — non-native speaker
phw — per hundred words
SLA — second language acquisition (discipline)
SR — speech rate
TL — target language
UP — unfilled pause
wpm — words per minute

*see also Appendix 4 which contains a list of tags used for error tagging LINDSEI_CZ

0. Introduction

Analysing advanced learner language is a truly fascinating enterprise – advanced learners appear to be extremely close to native-like performance and yet they are often recognizably unnative. Their language reflects complex cognitive processes of speech production (as described e.g. by Levelt (1989) and Kormos (2006) among others), it bears evidence of previous instruction and hints at what work remains to be done. It also offers unique material for SLA researchers concerned with questions of language transfer and the tempting perspective of cross-language comparisons of acquisition of the same language, and especially comparisons with native-speaker performance. How far are advanced learners from this traditional and natural goal? And what constitutes language proficiency? Since the 1980s there have appeared a number of influential models of language proficiency (e.g. Canale and Swain 1980; Canale 1983; Bachman 1990; Bialystok 1994; Bachman and Palmer 1996; Douglas 2000; Purpura 2004; Hulstijn 2015) and ways of describing it, e.g. competence-based models such as the Common European Framework of Reference (Council of Europe 2001). One of the latest but already very well established models of description is the complexity, accuracy and fluency (CAF) triad (Skehan 1996; Pallotti 2009; Housen et al. 2012) — an attempt to provide linguistic means for measuring performance and the underlying proficiency.

Whilst abroad there has been a sharp increase of studies of advanced-learner English, to date advanced learners of English whose first language is Czech have largely been ignored. And yet such research might have valuable implications both for the study of language acquisition and for pedagogy. If we knew more about the problems our advanced English learners experience, perhaps we would gain a deeper understanding of how they have — or should have — been taught and how to carry on teaching and learning. This is especially important for the students of English philology whose proficiency should reach the highest level imaginable. To find the answers we must look at the evidence at hand — learner language and its characteristics and dimensions.

The aim of my research is to describe the salient features of English produced by Czech advanced students of English philology. Whilst availing myself of the CAF methodology, I will narrow down my scope to analysing informal spoken production, and particularly its accuracy and

fluency. As regards accuracy, I will only deal with the description of errors (excluding errors in pronunciation). As for fluency, I will focus on its temporal aspects. I do not intend to analyse any of the subdimensions of complexity, and I do not provide a holistic view of fluency (Götz 2013).

The intended outcome of my study is manifold. The progress in the field of advanced-learner language analysis ought to be mentioned along with the demonstration of analytical techniques used in the process. The obtained results should enable comparison with similar studies of speakers of different L1s (e.g. Brand and Götz 2011; Götz 2013) and provide material for the study of universals in SLA. On a practical level, a deeper understanding of our learners should facilitate the design of syllabuses for practical language courses for students of English philology and by extension the design of any advanced English courses. As many of our students aim to become language teachers I cannot refrain from mentioning that I entirely embrace Coady and Huckin's (1997, 161) view of language teachers' proficiency which states that "One of their professional justifications is the proximity of their competence to that of native speakers." If our courses are to reflect this view we need to develop a deeper understanding of our learners' proficiency.

The thesis consists of 6 chapters. Chapters 1 provides the theoretical and contextual framework for my research. I discuss the CAF model on the background of theories of language proficiency, define accuracy and fluency and its components, including a historical overview of research in this field and a discussion of pedagogical implications of such debates. Chapters 2 and 3 deal with methodological issues. Chapter 2 describes the data set (a spoken corpus of c. 125,000 words) and the process of its compilation. Chapter 3 contains a description of the pilot study and of the techniques used for the error analysis and for obtaining fluency measurements. Chapters 4 and 5 present the results and their discussion. Chapter 4 describes the results of the study of accuracy, including a survey of frequency of errors, error rates, and an identification of persistent errors. The chapter provides an insight into the most problematic areas for advanced learners. Chapter 5 presents the results of the selected fluency measurements, namely speech rates, frequency of filled and unfilled pauses and the mean length of runs. These measures are pitched against measures of native-speaker performance analysed on a sample of a parallel corpus of approximately identical size and design.

Chapter 6 discusses the significance of the findings, their limitations, and their pedagogical and research implications. Tentative suggestions for future research are made. The thesis aims to make a contribution to the growing volume of learner corpus research studies and especially hopes to provide possible directions for the development of advanced-learner teaching strategies.

1. Language proficiency in the light of the complexity, accuracy and fluency model

This chapter deals with the complexity, accuracy and fluency model of L2 proficiency and performance. It describes the origins and the development of the model and the common definitions and ways of operationalisation of the individual components.

1.1 The CAF model

The concept of language proficiency has at its heart the question of what it is to know a language. For SLA researchers proficiency is the evidence of L2 development, for L2 teachers it is the state of the development which affects teaching practice, for L2 testers it is a gauge of achievement, and for the public (e.g. employers) it ought to be a guarantee of what an individual can achieve in a foreign language. For the language learner it is a combination of these views — it provides key information for further learning and it is linked to motivation. These different groups develop different views of the components of proficiency. Language teachers, students and testers are interested in practical language competence. For them, knowing a language is an ability or a skill which enables the learner to achieve practical goals in different contexts. This is traditionally linked to the model of the four skills (speaking, writing, reading and listening) and elements (phonology, orthography, grammar and lexis) (Lado 1961; Carroll 1968) and gives rise to competence-based models of proficiency which are typically based on descriptions of model performance in various situations and contexts. Such situations and the learner's performance in them are described and placed on a scale which has at its bottom cognitively simple tasks requiring simple language (e.g. asking for directions) and at its top cognitively complex tasks which require a vast range of linguistic means (e.g. delivering a lecture). Whilst competence-based models such as the Common European Framework of Reference (Council of Europe 2001) in their can-do statements discuss competence, the real target of their descriptions is performance. Competence is the underlying mental representation of language in the brain (R. Ellis 2008, 957) and as such it can hardly be measured.

Performance is the realization of competence which can only be translated into concrete language if the speaker is capable of it, if he is proficient enough. Proficiency is thus the ability to draw on and use competence in different tasks (Taylor 1986, 166; R. Ellis 2008, 976), and in this way it is closely linked to communication. The link between communication and proficiency was explored by Hymes (1972), who developed the concept of communicative competence as a counterpart and complementation of Chomsky's (1965) linguistic competence, and was further developed by Canale and Swain (1980) and Canale (1983) into a proficiency model which consisted of four competences (grammatical, socio-linguistic, discourse, and strategic) but which failed to show how these interacted. The model served as a foundation for Bachman's (1990) and Bachman and Palmer's (1996) concept of communicative language ability which described the interactions and deepened the understanding of strategic competence. It was based on the interaction of two basic components: organizational knowledge (the underlying principle here is grammatical control) and pragmatic knowledge (contextually appropriate production of meaning). The model spawned several variations and extensions which aimed to develop some of its details or add new dimensions such as applicability to language for specific purposes contexts (Douglas 2000), the role of background knowledge (ibid.), the process of meaning conveyance (Purpura 2004), the interaction between grammar and pragmatics (Chang 2004), the extension to specific skills such as writing (Hinkel 2002; Sasaki 2002) and reading (Weir 1997). All of the models mentioned so far recognize the multi-componential nature of language proficiency and the fact that the individual components may develop at different rates. They are especially useful for the fields of language pedagogy and language assessment as they identify areas of competence for which materials, classroom methodologies and tests can be developed.

In the field of SLA research measuring L2 proficiency is one of the key concerns, and especially the question to what extent it is feasible to measure proficiency using linguistic means. Three dimensions of proficiency are recognized as the key components of proficiency and performance. They are **complexity**, **accuracy** and **fluency**. The concepts were individually researched already in the 1960s, when especially the study of accuracy and error gave rise to dozens of studies (see Spillner 1991). The concepts were brought together in the context of the

communicative language classroom by Brumfit (1984) who distinguishes between fluency- and accuracy-oriented activities (see also Hammerly 1991). His theory is spurred by a conviction that a learner producing L2 cannot easily concentrate on all aspects of correct speech production and has to compromise one in favour of the other. Consequently Brumfit advises that these two components ought to be practised separately so that undue focus on accuracy does not detract from fluency and vice versa. Thus in fluency activities teachers are encouraged to condone decreased accuracy in favour of "getting the message across", in accuracy-oriented activities teachers are advised to accept slow, carefully planned responses. However sound this advice appears to be for classroom practice it failed to explain the inter-relation between the two concepts, neither did it attempt to initiate a discussion about the definition of accuracy and fluency. It was not until a decade later that Skehan (1996; 1998) proposed a model of proficiency which included not only accuracy and fluency but also complexity. The model (known as CAF) became very influential and opened a new field of SLA research.

Housen at al. (2012, 2) point out that it is since the 1990s that in countless studies CAF have featured as dependent variables, i.e. variables which reflect the effect of other factors on the production of language. These included a variety of factors such as task design (Skehan and Foster 1999; P. Robinson et al. 2009), planning (R. Ellis 2009; Ahmadian 2012), stay abroad (Trenchs-Parera 2009), learning contexts (Collentine 2004), age (Mora 2006) etc. In these studies CAF are seen as having a psychological, cognitive basis, they address the issue of the brain's capacity to process language in different conditions and contexts. They also invite attention to comparisons between native and non-native production (e.g. Götz 2013) but in this respect suffer from the dearth of native-speaker CAF data (Skehan 2009a).

The CAF provides a model of language performance in which the individual dimensions are accompanying features of cognitive and psycholinguistic processes. These depend on automaticity, parallel and controlled processing, proceduralisation, conscious awareness, use of attentional resources, type and speed of processing, difficulty and relative novelty of tasks, declarative and procedural knowledge, memory and retrieval (Levelt 1989; Kormos 2006; Sternberg et al. 2009). Skehan (1998) describes the interaction of the dimensions in his Limited Attentional Capacity Model

whose core principle is that speakers' information processing capacity is limited and as a result speakers have to choose to which of the dimensions they pay more attention. These decisions are commonly referred to as trade-offs, the two most notable ones being that between fluency and accuracy, and accuracy and complexity. Robinson's (2001a) Multiple Resources Model, on the contrary, claims that attention can spread to more dimensions at once and that the driving force is the complexity of the task whose high difficulty might demand high levels of linguistic accuracy and complexity but may compromise fluency. Housen et al. (2012, 6) point out that both models lack the support of empirical evidence.

1.2 The limitations of the CAF model

The CAF model is not without its problems. The dimensions are notoriously difficult to define, operationalize (see sections 1.3 and 1.4 below) and measure. This is largely because they are multi-faceted and multi-dimensional in themselves (Housen et al. 2012). As a result, studies purporting to deal with CAF frequently fail to provide accurate definitions and compared with each other they often concentrate on measuring different aspects and thus provide incomparable outcomes.

The studies of CAF leave one area blurred and that is the relation between performance and proficiency.[1] They use learner language as evidence and in this they deal with performance — the product and the process — and assume that performance is a reflection of proficiency. To an extent this is true: a highly complex, accurate and fluent production is determined by high proficiency but this may not work in the opposite direction — simple, inaccurate and dysfluent speech may not be

[1]A case in point is R. Ellis's (2009) study in which he copies a table transferred from Skehan and Foster (1999, 96–97) and inaccurately uses the column heading "definition" for what the two original authors labelled as "an outline of the three areas which compete for attentional resources while language is being produced". What they describe as areas, Ellis labels as definitions. The three dimensions' "definitions" are thus described as capacities, and in this way directly equalled with proficiency, which is confusing.

the result of low proficiency but may be caused by lack of concentration or interest, tiredness, excitement, illness etc.

Another limitation of the CAF model is that it does not subsume all dimensions of language production. Skehan (2009a) calls for the inclusion of measures of lexical use, whilst Fillmore (1979) in his discussion of the dimensions of fluency mentions for example the ability to have the right things to say, to be economical with language or to be creative and imaginative. Housen et al. (2012, 4) also mention acceptability and appropriateness as desirable extensions of the concept of accuracy. All of these, and possibly many more, could be valid and comparable dimensions of language production.

Despite its limitations, the CAF model presents a multitude of approaches to providing ways of describing linguistic performance and establishing links between performance and proficiency. I have tried to show the model's salient features and weaknesses. In the following sections I will describe the individual components in more detail.

1.3 Complexity

Housen and Kuiken (2009, 463) consider complexity to be "the most complex, ambiguous, and least understood dimension of the CAF triad". Part of the reason they give rests in the fact that two different types of complexity are recognized, namely cognitive and linguistic. The first sees complexity from the perspective of the learner. It is affected by learner variables (such as aptitude, L1 background, cognitive skills etc.) and refers to the difficulty the learner experiences when learning or producing L2 as subjectively perceived or as measured through psycholinguistic tests. Linguistic complexity, on the other hand, is independent of the learner and refers to the structural, semantic, functional and other intrinsic features of the L2 (Housen et al. 2012, 4). In the same volume Housen et al. use a working definition of complexity as "the ability to use a wide and varied range of sophisticated structures and vocabulary in the L2" (p. 2). Similarly, Ellis (R. Ellis 2003, 340) and Ellis and Barkhuizen (R. Ellis and Barkhuizen 2005) describe complexity as "the extent to which the language produced in performing a task is elaborate and varied." Skehan (2001) equates complexity

with elaborated language use and the learner's need to produce such language and accept the risks associated with it. Such a view is determined by Skehan's trade-off hypothesis according to which the learner prioritizes one dimension of his production at the expense of another as a result of being equipped with only limited capacity to process input and output (Meisel et al. 1981; VanPatten 1990; Skehan 1998). Skehan recognizes that it is the learner's choice whether he tries to produce simple or complex language.

Bulté and Housen (2012) attempt to design a "descriptive-analytic framework for future analyses of L2 complexity" (p. 26) based on a survey of complexity studies available to date. They distinguish between relative and absolute complexity. Relative complexity coincides with the concept of cognitive complexity (see above). As it is learner dependent it is highly subjective and may vary from learner to learner. Absolute complexity is a quantification of the number of elements of a given language features and the number of connections between them. It is further subdivided into discourse-interactional, propositional and linguistic complexity (also called global or system complexity). It is only the last one of this triad that receives separate attention. In its definition the authors speak of the learner's repertoire and describe it using such expressions as elaboration, size, breadth, width, richness, sophistication, number, range, variety and diversity (ibid., 25). They explicitly mention that these labels refer to various linguistic domains such as phonology, lexis, morphology and syntax and their subdomains, for example, inflectional/derivational morphological complexity or phrasal/clausal/sentential syntactic complexity. The proposed system is comprehensive and distinctly multi-faceted. It is, unfortunately, beyond the scope of the current chapter to describe it in detail. Some of the challenges it poses concern not only the operationalisation of the individual constructs, but also the difficulty and often the extreme laboriousness of obtaining measurements.[2]

[2]In a pilot study carried out for this thesis I carried out a segmentation of some of the interviews into AS-units (Foster et al. 2000). To this end I wrote a set of scripts which automated some of the tasks. Whilst the software made the process significantly more time effective, a full segmentation (i.e. the semi-automatic part and the manual part together) of one 15-minute interview still took an average of 5.5 hours to complete. With 50 interviews this would have amounted to 275 hours. The

Measures of grammatical complexity express, for example, the mean length of selected unit (e.g. turn, utterance, clause, T-unit, c-unit, AS-unit), level of subordination (expressed as a ratio between e.g. clauses and a selected type of unit), frequency of a specific feature (e.g. relative or other subordinate clauses, verb arguments), syntactic sophistication (e.g. frequency of passives, conditionals, infinitival phrases), or morphological sophistication (e.g. variety of tenses, measures of affixation). Lexical complexity includes measures of diversity (e.g. type-token ratio, Guiraud Index), density (ratio of lexical words and function/total words), and sophistication (use of low-frequency words).

A question arises whether complexity and its subdimensions is a suitable concept for the description of spoken language. Housen and Kuiken (2009) associate complexity with the production of syntactic embedding and subordinate clauses in relation to the number of clauses produced. Such an understanding of complexity goes against the nature of spoken language which uses less complex phrases than written language and which prefers coordination or simple parataxis to grammatical subordination (Miller and Weinert 2009). In this light, using complexity as one of the dimensions of informal L2 speech performance seems to be a fallacy. Moreover, complexity is linked to register — a formal academic discourse will demand higher levels of complexity than informal small-talk. The competent speaker will be able to judge when to use complex language and when not and he will have the flexibility to do so. Complexity seems to differ from accuracy and fluency in that it appears to be an arbitrary component of language production — the speaker may "turn it on or off" according to the personal or situational requirements. There are many situations in which high complexity would be impedimental, inappropriate or even ludicrous. Accuracy and fluency, on the other hand, seem to be intrinsic features of spoken production — a speaker does not choose (at least not usually) to be accurate or inaccurate, fluent or dysfluent. Accuracy and fluency are true epiphenomena of language production, whilst complexity is often simply an option.

---

segmentation itself is, however, only a preparatory stage for performing some of the complexity analyses.

As in my thesis I deal with the description of advanced informal spoken learner language and the interviewed speakers were not explicitly asked to pay attention to using complex language I have decided to abstain from measuring the complexity of their output.

1.4 Fluency

It has become something of a cliché to start a chapter on a linguistic phenomenon by saying how difficult it is to define it. This is especially true when the phenomenon in question has a life of its own outside linguistics. Examples might include the word *word* itself, and indeed any of the many expressions which exist as part of our everyday word stock starting with simple concepts such as sound, pause or sentence and going all the way to complex ones such as competence and proficiency. The word *fluency* is a perfect example. It has at its root the trace of an Indo-European morpheme *bhleu with the meaning *to swell, to overflow* and as such it entered into many languages. In Czech we have *plynulost*, in Italian, French and Spanish *fluidità*, *fluidité, fluidez*, in German *flüssigkeit* and we could carry on. The connotation of these words is one of quantity in effortless motion. A similar concept is often expressed in various languages with reference to speed or swiftness as in the Czech *zběhlost*, Italian *scorrevolezza* (from *correre* to run), Russian беглость, or Polish *biegłość*. These words do not necessarily refer only to speech — we can read fluently, but we can also be fluent at dicing onions or driving a car. At the receiving end we can also speak of fluency in understanding, i.e. receiving and interpreting the linguistic code at the speed at which we are encountering it. When we call somebody a *fluent speaker of a language,* however, we mean even more than just the effortless production of a stream of language, we refer to the underlying competence or proficiency. Thus in lay terms fluency is not just a dimension of performance, but an evaluation of one's linguistic capacity or global proficiency (Koponen and Riggenbach 2000). We also tend to associate fluency with native-speaker production but we do not speak of natives as of fluent speakers, the term is reserved for L2 speakers.

In linguistics, fluency has been known to mean different things to different people. In the following paragraphs I will describe some of the key approaches to its conceptualization. I fully

realize that fluency relates to reading, writing and listening but henceforth I will concentrate only on fluency in speech.

Fillmore's (1979) seminal essay *On Fluency* describes four dimensions of fluency. The first is "the ability to fill time with talk" and refers to the effortlessness and quantity mentioned above. The second is "the ability to talk in coherent, reasoned, and 'semantically dense' sentences". In praising the avoidance of producing "semantically empty material" (ibid.) it resembles Grice's maxims. The third type, "the ability to have appropriate things to say in a wide range of contexts", is a socio-pragmatic skill. The fourth type, "the ability [...] to be creative and imaginative" relies on the capacity to look ahead and make the "most sonorous or clever" selections. This dimension is closely related not only to language use but also to personal characteristics. The real (and generally ignored) contribution of Fillmore's article is his suggestion of research possibilities, where he does not only call for the operationalisation of fluency but he also suggests that quantitative measurements (such as speech rate) ought to be supplemented by qualitative evaluations by panels of reliable judges (see Lennon (1990b) and Götz (2013) for examples of such a study).

Brumfit (1984) presents a pedagogical view of fluency. He defines it as "the maximally effective operation of the language system so far acquired by the student" (quoted in Riggenbach 2000, 69) and recommends separating fluency-oriented classroom activities from accuracy-oriented ones. This has become a mainstream idea for the design of Communicative Language Teaching methodology. Lennon's (1990a) seminal study of fluency has a pedagogical background as well. He sets out to explore fluency quantitatively in order to find measures which could be used as a gauge of progress or as a goal to be achieved by learners. He establishes the distinction between a broad sense of fluency, i.e. global oral proficiency, and a narrow sense[3] which includes only some components or their combination, typically temporal phenomena (e.g. speech rate, pausing, length of runs) (cf. Schmidt's (1992) view of fluency as a primarily temporal phenomenon) or hesitation phenomena

[3]In a later study (Lennon 2000) he refers to the two types as higher-order and lower-order fluency. The latter includes temporal and also perceived fluency which are grouped together as both of them can be measured instrumentally or impressionistically (p. 25).

(e.g. repeats, self-corrections, and filled pauses) and hopes that one day a machine analysis of spoken language fluency might become possible. In a later study (2000, 26) he defines fluency as "the rapid, smooth, accurate, lucid and efficient translation of thought or communicative intention into language under the temporal constraints of on-line processing" highlighting the multi-faceted nature of the phenomenon. Skehan (2003) and Tavakoli and Skehan (2005) distinguish between three subdimensions: speed fluency (speech rate), breakdown fluency (count, length and placement of pauses) and repair fluency (self-corrections, false starts, repeats, misformulations).

Skehan's (2009b) view of fluency as "the capacity to produce speech at **normal** rate and without interruption", and Ellis and Barkhuizen's (2005) "the production of language in real time without **undue** pausing or hesitation" (my highlighting) raise the notion of norm. As language users we tend to notice when somebody speaks too slowly or too fast or whether he uses dysfluencies frequently even though there are no explicit norms thereof. We appear to be in possession of an automatic evaluative system (this might coincide with the processes we use to monitor our own speech) and the question arises whether in judging L2 speech (e.g. as language examiners) we apply second or first language parameters. Fluency is, nevertheless, clearly not only an objective, quantifiable dimension, but also an impression made on the hearer.

Segalowitz (2010) distinguishes three facets of narrow fluency: cognitive (the smoothness of the underlying processes), utterance (acoustically measurable aspects of performance) and perceived (the speaker's fluency impression on the hearer). Similarly Götz (2013), who does not directly define fluency but offers a holistic view of it, mentions three abstract categories: productive, perceptive and non-verbal fluency. Productive fluency is performance-based and is made up of a combination of temporal variables, formulaic sequences and fluency enhancement strategies (e.g. repeats, filled pauses etc.). Perceptive fluency takes into account the effect speech has on the listener and includes such dimensions as accuracy, idiomaticity, intonation, accent, pragmatic features, lexical diversity, register and sentence structure. Non-verbal fluency is characterized by the use of paralinguistic features which accompany speech such as gestures, facial expressions, body language, looks and emblems. Whilst this is supposed to be an attempt to provide a "framework [...] for thinking systematically about

fluency" (p. 10) it is more of a catalogue of features typically or more loosely associated with the phenomenon without attempting to describe how they interrelate and overlap.

1.4.1 The sources of fluency

Whilst the majority of fluency research concentrates on measuring its temporal aspects there is a growing body of research which taps into its cognitive foundations. To understand these it is essential to start with an overview of a model of speech production. The most widely accepted and most frequently cited one is Levelt's (1999) modular model. Levelt assigns the production of speech to a number of autonomous components. Their autonomousness is crucial in that it makes it possible for the many associated processes to happen simultaneously.

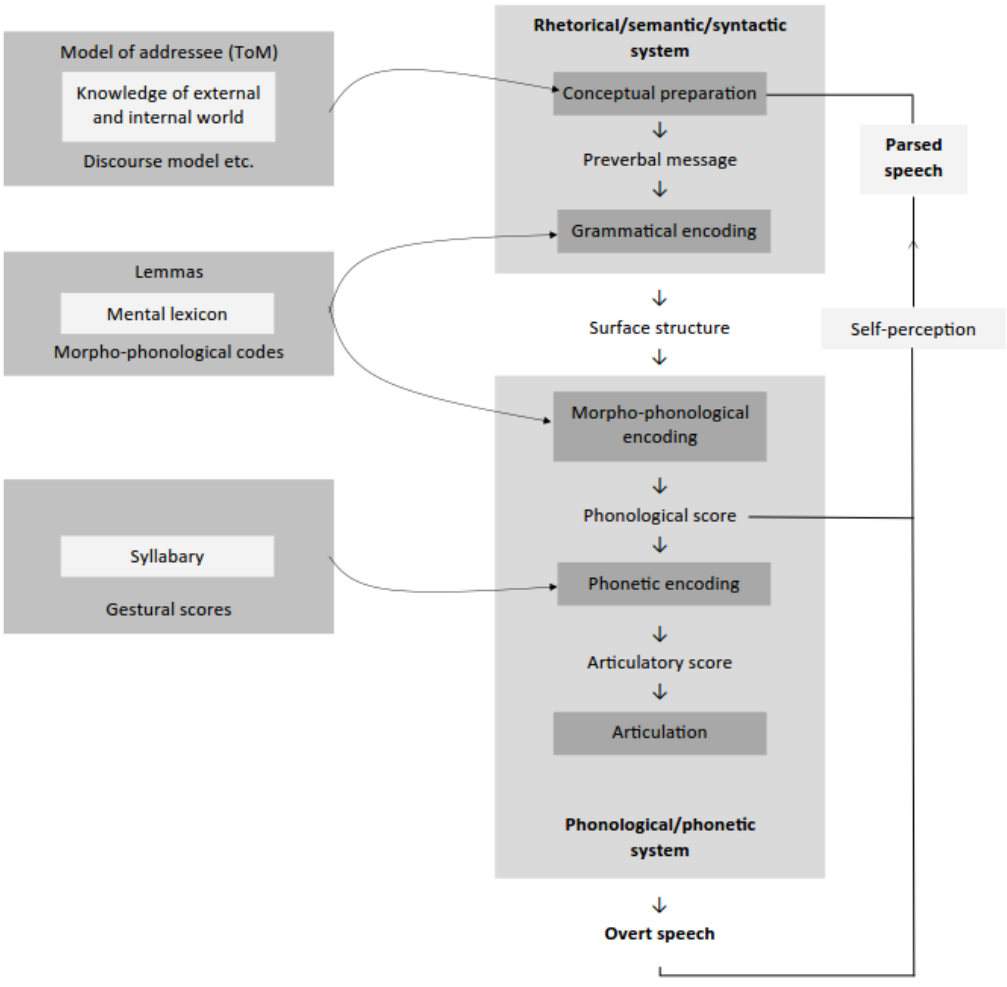

Fig. 1–1 Levelt's (1999) blueprint of the speaker's mind (Kormos 2006)

As shown in Fig. 1–1 there are two cardinal systems: the rhetorical/semantic/ syntactic system and the phonological system. They draw on three knowledge stores: the mental lexicon, the syllabary, and the general world knowledge store. Speech production starts by the conceptualization of the message which is subsequently encoded and articulated. A synchronous system of monitor modules guarantees the desired progress by inspecting that the individual stages fit the speaker's intentions. In more detail, the mechanism is as follows. During the initial stage the speaker generates the concept and by means of macroplanning and microplanning determines the communicative intention (speech act), the selection and order of the information, the propositional content of the message, the relation to what is new and old, temporal and distal relations. This produces a preverbal plan which can now be grammatically encoded. This affects the selection of lexis and appropriate syntax. The mental lexicon is accessed during this stage as it contains lexical entries: lemmas (i.e. syntactic parameters of the lexical entry) and lexemes (i.e. morpho-phonological information). The result is a surface structure which is passed onto the next stage, the morpho-phonological encoding. Here the morphemes are selected, the stress and pitch are determined and this is all assigned to appropriate phonemes. The process produces a phonological score (so called internal speech) which provides the basis of the production of the articulatory score which draws from the syllabary (the store of articulatory gestures). The whole process is completed by articulation, the conversion of the articulatory score into speech.

Levelt's model, which does not account for production in L2, served as a basis for models of bilingual production, most notoriously de Bot's (1992) and most recently Kormos's (2006). The fact that Kormos's model is considerably similar to Levelt's shows that L1 and L2 speech processing are similar, which is confirmed by a number of neuroimaging studies (reviews in Abutalebi et al. 2001, 2005, quoted in Kormos 2006). Kormos first of all considers the automaticity of Levelt's model, which is due to parallel processes. Cognitive psychology claims that parallel processing is facilitated by the availability of attentional resources. If an event can happen with little or no conscious awareness, attentional resources are available for other processes which can then happen in parallel (e.g. simultaneous reading and scratching one's head) (Sternberg et al. 2009). If, however, a task

requires attention, attentional resources are more easily depleted and processing happens only sequentially. Should we accept that this hypothesis applies to Levelt's model, it would explain why L2 speech is less fluent and why interruptions such as filled and unfilled pauses happen at places where native speakers do not generally make them (cf. Kjellmer 2003). As Kormos points out (p. 166) it would also account for the development of fluency with increasing proficiency. Central to Kormos's model (see Fig. 1–2) is a global memory store (placed in long-term memory) which is made up of episodic memory, the lexicon, the syllabary, and a store for declarative knowledge of L2 rules.

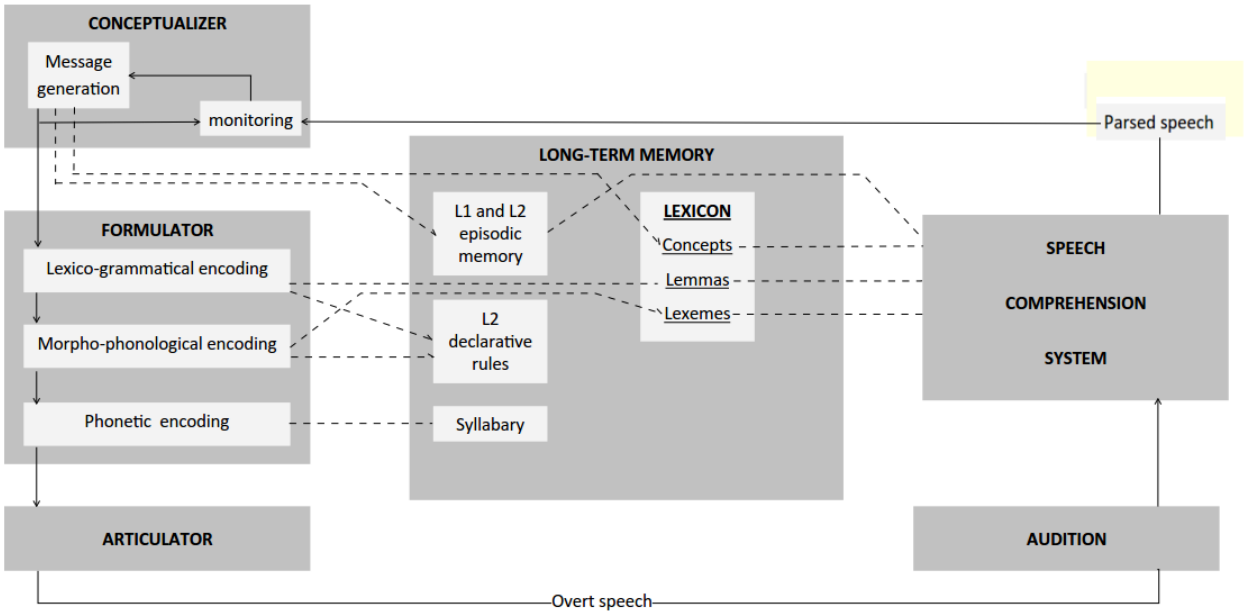

Fig. 1–2 Kormos's (2006) model of L2 speech production (Kormos 2006)

Whereas episodic memory, lexicon and the syllabary are shared between L1 and L2, the L2 system adds the store for L2 rules, which contains such rules that have not been automatized yet.[4] It is at this stage that a missing or insufficiently automatized L2 rule is substituted by a communication strategy which might involve, for example, the borrowing of a rule from L1 (resulting in a transfer error). Monitoring happens at all stages but it uses up attentional resources and the speaker has to make a choice as to whether to pay more attention to form, lexis or grammar (cf. Skehan's trade-off hypothesis).

[4]In L1, rules are thought to be automatic, and part of the encoding systems.

As has been explained, in both models much depends on the automaticity and synchronicity of the processes. Once each module completes a stage, it passes on information to the next module and thus becomes instantly available for subsequent processing. It is in the multi-tasking nature of the model that fluency is guaranteed.

Dörnyei (2009) explains fluency as a skill and describes its acquisition using a theory of skill-learning which is assumed to have three stages: cognitive, associative and autonomous (sometimes these are called *declarative*, *procedural* and *automatic*). During the cognitive stage, instructions and explanations are provided as a form of declarative knowledge aiding the first-time performance and subsequent practice attempts. During the associative stage, declarative knowledge is transformed into procedural knowledge in a process called proceduralisation. An alternative explanation sees the process not as one of transformation but of building a parallel structure. During the automatic stage the learner learns to spend fewer attentional resources on the process which he can now perform without conscious awareness (Sternberg et al. 2009). Fluency is thus a result of automatization which is closely related to practice. Dörnyei (2009, 163) warns that highly automatized skills are strongly context-specific which might explain the high occurrence of variance in learners' performance.

Lennon (1990a, 389) argues that fluency is purely a performance phenomenon for which there is no "fluency store", and in this way he sees it as distinctly separate from forms of linguistic knowledge (e.g. of grammar rules or of lexis). Such a view ignores the value of linguistic knowledge of which fluency is an epiphenomenon and without which very little language production would be possible.

Fluency, as a continuous production of meaningful and rule-abiding speech, is often explained in terms of automatised procedural linguistic knowledge and the knowledge of formulaic speech, speech chunks, or lexicalized sentence stems (Pawley and Syder 1983; R. Schmidt 1992; Towell et al. 1996). Skehan (1998) claims that two systems contribute to the fluent production of language: a memory-based one which stores formulaic chunks as prefabricated exemplars which are accessible as wholes (cf. Sinclair's idiom principle) and a rule-based one which allows the computation of novel utterances and the combination of chunks into larger stretches of language. This dual system

is characterized by high permeability with the speaker constantly either switching from one to the other or using both simultaneously. The implication of such a system would be the existence of a vast store of memorized exemplars which would be larger in L1 than in L2 (cf. Foster 2001). This would explain why native speakers' speech rate is generally significantly higher than that of L2 learners however advanced.

Besides the cognitive underpinnings, fluency is affected by external factors (Housen et al. 2012). These may include a variety of learner variables and personality traits especially temperament, extraversion (Busch 1982; van Daele et al. 2006), psychoticism, neuroticism or anxiety, the speakers' L1 speech habits (Hincks 2008; Derwing et al. 2009), and also the fact whether the speaker has something to say. The precise effect of any such external factors is yet to be further researched.

1.4.2 Operationalisation of fluency

Fluency is as difficult to operationalize as it is to define. The key question is the validity of measurements and whether empirically attested fluency would coincide with the impression the speaker makes on the hearer, the so called perceptive fluency. However difficult operationalisation might be, the majority of empirical studies seem to coincide in the selection of features to be observed. At least as far as temporal or productive (Götz 2013) fluency are concerned. It was Lennon (1990a) who, following Möhle's (1984) suggestion that the assessment of fluency might be linked to certain temporal features, compiled one of the first catalogues of quantifiable fluency-related measures. He did so in a study of German advanced speakers of English whose fluency was first measured by the author of the study and then impressionistically evaluated by a panel of ten judges. Two tests were carried out — one before and one after a six-month stay in England. Lennon included speech rate which was measured as the number of pruned or unpruned[5] words per minute. Then he included

[5]Unpruned words are all words produced, whilst pruned words do not include repeats and self-corrections.

repetitions[6], self-corrections and filled pauses establishing their relative frequency (expressed as rate per T-unit)[7] and the ratio between pruned and unpruned words. Next he considered the ratio between speech time and time taken up by unfilled and filled pauses, total and mean pause time at T-unit boundaries, percentage of T-units followed by a pause, and the mean length of speech runs (the amount of text between pauses).

These measures could be divided according to the key areas of focus into the following groups: speech rate, measures of silence, length of runs, hesitation phenomena, repair phenomena, and use of formulaic language.

1.4.2.1 Speech rate

Speech rate appears to be the most robust dimension and as shown by Kormos and Dénes (2004) and Bosker et al. (2013) it is a strong predictor of perceived fluency. In the context of SLA the most frequently used measure is the number of words per minute (wpm), which is obtained by dividing the total number of words (pruned or unpruned) by the total speech time in minutes including pauses. More accurate picture would be provided by counting syllables per minute (as especially more advanced learners might use longer words) but this is much more laborious to count. Besides some studies (e.g. Griffiths 1991) suggest that the variance between learners is low and the average syllable count is 1.15 syllables per word. Thus the syllable rate would be simply 1.15 times higher than the word rate, and consequently not providing any greater precision. This is due to the fact that spoken language is characterized by the use of shorter words. Typically, studies do not work with a definition of a *word* and we can only assume that a graphic word is meant (i.e. delimited by spaces on either side in the transcript). Consequently we do not know whether contracted words are counted as one or two words. In her 1984 study Möhle worked with the variable articulation rate,

[6]Lennon distinguishes between words repeated for rhetorical effect and those repeated as a result of planning problems. However, he calls both types repetition, whilst later research labels the first as repetitions and the latter as repeats.

[7]Defined by Hunt (1970) as one main clause and all its attendant subordinate clauses and non-clausal units.

which she calculated as syllables per minute minus silent pauses, that is the time actually spent producing sound. Cucchiarini (2002) counted the number of phonemes per time unit but the question arises whether such subtle distinction provides a utilisable measure in research. The main advantage of using words per minute is its user-friendliness — it is easy to count in the transcript and easy to imagine when one is confronted with measurements. The unit is frequently used in studies dealing with reading speed as well.

Speech rate is known to vary to a great degree. In a survey of studies Götz (2013, 15) found that the speech rate of native speakers in conversation is given as a range of 120 to 260 wpm. Different communicative situations researched provide different typical rates. There are genre requirements as well — an official speech addressed to a large audience is likely to be slower than a news report. In a complex study of speech rate in British English, Tauroza and Allison (1990, 102) give the following ranges (Table 1–1).

Table 1–1 Range of speech rates by words per minute for four different types of speech (Tauroza and Allison 1990, 102)

	Radio	**Lecture**	**Interview**	**Conversation**
Faster than normal	190	185	250	260
Moderately fast	170—190	160—185	210—250	230—260
Average	150—170	125—160	160—210	190—230
Moderately slow	130—150	100—125	120—160	160—190
Slower than normal (below)	130	100	120	160

Speech rate is also likely to vary within one conversation depending on the complexity of the topics discussed. Studies also show, that various non-linguistic factors are at play such as gender (Whiteside (1996) shows that men are faster speakers than women), age (Ramig (1983) documents the decrease of speech rate with age) and emotional state and the stress level (Hausner 1987). Speech rate in a L2 is reduced compared to one's mother tongue as demonstrated by Hincks (2008) who reports a difference of 23% in a contrastive study of Swedish advanced learners of English delivering a presentation in English and in Swedish. Nevertheless other studies (Segalowitz and Freed 2004; Derwing et al. 2009) are cautious to put a direct link between L1 and L2 speech rates and other fluency

measures as there are too many confounding variables which are impossible to control for (e.g the amount of exposure to the L2, the level of proficiency, or language aptitude).

The effect of the speaker's speech rate on his audience may be of various nature as well. Lennon (1990b) showed that an increase in speech rate during a stay abroad resulted in a higher fluency rating awarded by the judges but in a later study (2000) he shows that speaking faster does not automatically increase one's status as a more fluent speaker. Munro and Derwing (2001) warn that increased speech rate is not always an advantage as they suspect that at higher rates "L2 speakers may be prone to making more segmental and prosodic errors". They also point out that when more heavily accented speech is delivered at a faster rate it is harder to understand.

The majority of L2 studies dealing with fluency consider speech rate as its principal component. They are especially of longitudinal nature, describing speech rate development over a period of time (e.g. Towell 1987; Towell et al. 1996) and typically as a result of a stay abroad (e.g. Lennon 1990; Freed 1995). The documented speech rate increase proves that speech rate is closely linked to proficiency and its growth during an extended period of study and especially exposure and concomitant practice typical of stay-abroad contexts. Ahmadian (2012), for whom speech rate is such a robust measure of fluency that he does not include any other, studies the effect of planning on CAF and concludes that speech rate is linked to spontaneity and is adversely affected by careful planning. Pressured, online planning, on the other hand, results in an increase in speech rate possibly as a result of a greater reliance on the use of prefabricated language and implicit knowledge (p. 145). Brand and Götz (2011) and Götz (2013) compare native speakers and advanced learners and show that the latter are not only slower but also more homogeneous as a group, as there appears a greater dispersion amongst the native speakers' measures.

1.4.2.2 Measures of silence — silent pause phenomena

Unfilled pauses are a natural phenomenon occurring in native English at a frequency of over 19,000 instances per million words (Biber et al. 1999, 1054). Fillmore (1979) and Lennon (1990a) note the multifunctionality of pauses: they can be used for rhetorical, stylistic, physiological (taking

breath) and speech-planning purposes. A pause can also be a reaction to an external stimulus (e.g. following a distraction), or the result of the speaker's being preoccupied with other thoughts. Griffiths (1991) claims that pauses facilitate comprehension as they provide both processing time and structure. Chambers (1997) considers intrasentential pausing patterns and observes that pausing within an unfinished utterance is a sign of increased planning pressure as a result of searching either for what to say (idea) or for how to say it (choice of grammar and/or lexis). Whilst the reasons for pauses are difficult to explain with any degree of certainty, unfilled (also called silent) pauses are traditionally seen as markers of fluency (or dysfluency) and have been the focus of many studies. These concentrate especially on their frequency, duration, distribution and location, but also on their effect on speech rate.

Riggenbach (1991) and Freed (1995) provide empirical evidence of the relation between fluency and pause frequency with users judged as less fluent using longer and more frequent pauses. Cucchiarini (2002), however, finds that the length of pausing has almost no effect on perceived fluency ratings (cf. also Wood 2012).

Chambers (1997) distinguishes between natural pauses (e.g. those occurring at structural junctures or after a completed semantic unit, and unnatural ones (placed elsewhere) and Lennon (1984) speculates that the distribution of pauses differs in native and non-native speech. Thus, for example, Bada (2006) finds in her study of the placement of pauses preceding and following 'that' a different pattern of placement amongst native and non-native speakers. Pawley and Syder (1983) claim that native speakers pause or slow down typically at or near clause boundaries and only rarely in the middle of clauses.

Riggenbach (1991) suggests that especially in less fluent speech unfilled pauses occur in combination with other markers of dysfluency. She calls such combinations "disfluency chunks" and shows that these occur less frequently with speakers perceived as more fluent.

One way of looking at silence is comparing the amount of silence to the amount of speech produced. This is referred to as the phonation/time ratio and calculated as a ratio of speech time and total time of the speech sample (Towell 2002). It provides a rather crude measure which is hard to

interpret as it provides no indication as to the location and explanation of the pauses used. Moreover it relies on an accurate measuring of all pauses which is a laborious and time-consuming procedure.

1.4.2.3 Mean length of runs (MLR)

The higher the frequency of unfilled pauses, the more fragmented speech becomes. Pauses mark the natural boundaries of the so-called speech runs, which are defined as the speech occurring between two pauses. A mean length of runs (calculated as a sum of all lengths of runs divided by the number of runs) is a frequent fluency measure which is an expression of the degree of fragmentation. The longer the runs, the more control the speaker has over his performance. L2 speakers typically produce shorter runs than native speakers (Raupach 1980) and as can be expected the MLR is not only an expression of fluency but also of proficiency — Towell (1987), Towel et al. (1996) and Lennon (1990b) report significant increases in MLR in their subjects over periods of stay abroad. The MLR is thought to be linked to the use formulaic chunks which facilitate planning during speech (Wood 2012, Götz 2013). Such a hypothesis would explain the growth in MLR with developing proficiency — as speakers progress they develop a larger repertoire and consequently are capable of producing longer stretches of text. However, we do not know precisely what the speaker does and how far ahead he looks when producing a lexical chunk. Also, identifying lexical chunks presents considerable problems.

In her cross-linguistic study of fifty native and fifty non-native advanced speakers Götz (2013) finds the L2 utterances significantly shorter. She also finds an unexpected disparity between the width of dispersion of the native and non-native data. The natives' MLRs average at 12.68 words and range from c. 8 to 19 words, while the non-natives' MLRs average at 5.88 words and range from c. 3.5 to 9 words. A question arises whether such a large variance is not an outcome of inaccurate measuring. Nowhere does the author tell us whether the same criteria for determining the runs' cut-offs were applied to both sets of the data. Götz explains the low MLR of her non-natives by the very high frequency of unfilled pauses — these are, however, not given in concrete numbers for either of the groups. A detailed comparison is thus not possible.

Working out the MLR is a laborious process. Sound parameters must be specified for the choice of runs for inclusion,[8] and especially for the length of pauses which mark the runs' boundaries. These then have to be measured. Götz reports identifying 200 runs for one 15-minute interview. With 100 interviews this involves 20,000 runs. A simpler and less time-consuming procedure ought to be found. It would appear that a pause rate, i.e. the number of pauses per hundred words, might provide some indication as a higher frequency of pauses ought to predict a higher level of fragmentation. This would, however, have to be empirically tested.

1.4.2.4 Repair phenomena

Spontaneous production of speech is characterized by the presence of the so-called repair phenomena (Wood 2012, cf. Skehan's (2003) repair fluency), performance phenomena (Götz 2013) or speech management strategies (Rühlemann 2006; Götz 2013). They include self-corrections, repeats and false starts. All of these signal processing problems. The existence of self-corrections and false starts attest Levelt's (1987, 1999) monitoring processes. Speech production happens in parallel with self-perception and the speaker's monitoring system evaluates whether the outcome matches the intention. This is one of Biber et al.'s (1999, 1066) principles of online production ("qualification of what has been said"). Self-corrections are also a direct evidence of Kormos's (2006) store for declarative knowledge of L2 rules as these are drawn upon during the monitoring process. Self-corrections (Biber et al. call them retrace-and-repair sequences) may involve grammatical, lexical or phonological choices and should be labelled as such only when they involve a correction of an error. In other cases, or if this is impossible to determine, it is more accurate to talk about reformulations or false starts. These are also a reflection of the monitoring processes but of more complex nature. The speaker might reformulate for a number of reasons such as the need for more precision, the subsequent evaluation of the perlocutionary act, as a face-saving act, or in reaction to any "needs arising from the interactive nature of real-time conversation" (Rühlemann 2006, 402). Self-corrections

[8]Götz (2013) excludes short answers, interrupted utterances and inaudible passages.

have been found to have little effect on perceived fluency (Lennon 1990b, Riggenbach 1991, Freed 1995, Bosker et al. 2013).

False starts are similar to self-corrections in that they involve a sudden interruption of the utterance. Whilst self-corrections are retraced and restarted, false starts are typically abandoned. Interruptions may be explained by a variety of reasons including external interruptions (by another speaker or event) or a change of communicative intention (e.g. the speaker changes his mind). False starts and self-corrections may be difficult to distinguish from each other.

Repeats present a distinctly different phenomenon. They are not linked to the monitoring processes but are rather epiphenomena of planning. It is a strategy which on the one hand complies with the principle of keeping the floor, or Biber et al.'s "keep talking" principle, and on the other it buys time for planning. Typically, function words are more prone to being repeated than lexical words. Repeats most frequently, but not exclusively, occur at the beginning of clauses as this is where planning pressure is at its peak.

Although repeats are a natural component of speech and significantly ease planning pressures, L2 speakers have been found to underuse them (especially articles and determiners) (Götz 2007; 2013). They also have a different distribution from native speakers and use repeats more frequently also within clauses. In L2 speech, repeats are often combined with other hesitation phenomena, especially filled and unfilled pauses, creating dysfluency chunks. Carter and McCarthy (2006, 173) advise that "repeats are not to be taken as a sign of sloppy or lazy performance" but as an effective device for maintaining fluency.

In calling these phenomena speech management strategies (Rühlemann 2006, Götz 2013) we acknowledge that in spontaneous speech they are not perceived as marks of dysfluency but as highly natural and functional components. Götz (2013) even suggests that teaching such strategies to L2 learners might prove beneficial to their fluency gains.

1.4.2.5 Filled pauses

Filled pauses (FPs), also referred to as fillers, are possibly the most characteristic feature of spoken discourse. They vary[9] in length, pitch, frequency, presence of nasalisation, and distribution and in their idiosyncraticity characterize not only spoken discourse, but also the speaker and the genre of speech. Along with unfilled pauses, they are considered by many to be typical markers of dysfluency. Recent research suggests their role in the production of fluent speech ought to be reconsidered (Christenfeld 1995; Clark and Fox Tree 2002; Kjellmer 2003; O'Connell and Kowal 2004; Rühlemann 2006; Tottie 2012; Götz 2013). They are similar to repeats in terms of their distribution and consequently Götz (2013, 36) suggests that they possibly function in a similar way, for utterance planning purposes (see also Tottie and Svalduz (2009) who call them planners). However, speakers frequently use them also to indicate that they intend to carry on talking, and do not wish to be interrupted, in which case they are placed at the onset of a dependent or subordinate clause (Biber et al. 1999, 1054). It is in their placement and distribution that differences in native and non-native speaker use are to be found. In a corpus-based study Kjellmer (2003) shows that FPs introduce thought units at word, phrase and clause levels, and that they function as hesitation markers, speaker-turn signposting markers, highlighters, introductions to correction, and attention drawing devices. He advocates their use as fully natural as they are "most of the time guiding and lubricating elements that facilitate communication" (p. 191). Mora (2006) sees their function in maintaining the appearance of fluency and sees them as part of *compensatory fluency*. Clark and Fox Tree (2002) understand them as convenient "solutions to problems in speaking" and propose that they are to be counted as "conventional English words. Speakers plan for, formulate and produce them just as they would any word." (p. 73).

However natural the use of FPs might be, they may at times be fairly audible (e.g. speaker CZ008 in LINDSEI_CZ) and may even disturb or irritate the listener. Christenfeld (1995), however, tried replacing filled pauses with silent ones and discovered that listeners preferred the original,

[9]Common transcription forms include ah, eh, er, uh, um, erm, and mm.

unedited version. He concluded that it is people concerned with the speaker's style who are more sensitive to their use, whereas listeners attending to substance either fail to notice them or they consider the speaker less anxious. In our evaluation of spoken performance we are affected by principles of well-formedness which are based on written language (O'Connell and Kowal 2004; Linell 2005), in which filled pauses are not used even to such an extent than in literary transcriptions of dialogues they are largely avoided by the authors. O'Connell and Kowal (2004, 461) state the frequency of 2.5 FPs in four novels they analysed as opposed to 27 FPs in one minute of spontaneous spoken discourse. What would present a huge obstacle in writing may go virtually unnoticed in normal speech.

Taking into account all that has just been said it is clear that using FPs as one of the measures of fluency is problematic. Not least because problems may be encountered when identifying FPs and distinguishing them from prolonged syllables (e.g. *He is going to:. mention it*) or, in English, from the indefinite article. Fluency studies work, for example, with absolute and relative frequencies and also with their location. Lennon (1990a) operates with total FP-time as a percent of total delivery time, and also considers the location and duration of FPs with regard to the T-units. Mora (2006) explores the frequency of "clause-internal filled pauses" but unfortunately ignores the fact that clauses are not suitable units of segmentation of speech (Foster et al. 2000). Götz (2013) compares native and non-native rate of FPs per hundred words. The studies document differences between native and non-native use, the effect of stay abroad and that of growing proficiency. The pedagogical implications of these studies are purely theoretical and have not been empirically verified.

1.4.2.6 Discourse markers and small words

Besides the use of FPs, the impression of speech as fluent has been shown to be contributed to by the use of discourse markers and smallwords. Hasselgren (2002) includes discourse markers such as *well, right, all right, okay, you know, you see, I know, I see, oh, ah, I think, I mean, like, sort of/kind of, a bit, just, or something, and everything/and that/and stuff/and things, not really.* She shows that more fluent users (as measured using temporal features) have a more native-like frequency and

distribution of smallwords and suggests that smallwords may facilitate speech. Her non-native speakers, however, used a much smaller range of smallwords and showed a greater propensity to overuse some of them (she calls these "lexical teddy bears", p. 155). She argues that "the acquisition of smallwords is a crucial step in the attainment of native-like fluency".

Studies of L2 use of smallwords and discourse markers (De Cock 2000; Hasselgren 2002; Müller 2005; Mukherjee 2009; Götz 2013) typically show their underuse compared to L1 production. Combined with Hasselgren's (2002) findings about the effect of smallwords on fluency measures this should have pedagogical implications for both testing and teaching. However, empirical research on the efficacy of related instructional procedures ought to be carried out in this area.

1.4.2.7 Conclusion to operationalisation of fluency

As is apparent from the previous list of features which are commonly measured as possible dimensions of fluency, the phenomenon is clearly multi-faceted and rather unwieldy. We cannot be entirely sure that all of its dimensions contribute to it in the same way, and that the whole is the sum of its parts (Fulcher 1996). It is also likely that some of the measures overlap and different ones measure the same thing (e.g. a high frequency of pauses undoubtedly results in shorter lengths of runs). Operationalisation is complicated precisely because of the multifacetedness of the whole construct, which has an undesirable effect in that too many notions are grouped under the same category and in effect provide a more blurred picture rather than a wider one (Pallotti 2009, 599). Skehan (2009b) suggests that "fluency needs to be rethought if it is to be measured effectively" and calls for a greater sophistication in its measurement (e.g. individual evaluation of pauses to determine whether they are rhetorical or dysfluent). Besides the highly problematic operationalisation, one of the most important drawbacks of measuring fluency is its extreme laboriousness — transcriptions have to be evaluated along with the recordings, which is highly time-consuming.

Since the early days of fluency research studies have tried to verify the validity of the measures by carrying out perceptive evaluations. This has confirmed that raters' evaluations are affected not only by temporal features and dysfluencies, but also by accuracy, intonation,

idiomaticity, accent, lexical diversity, pragmatic features, sentence structure, rhythm, confidence in speech (Freed 2000; Götz 2013), which all contribute to the general impression of native-likeness. Inter-rater reliability in these studies shows that the perception of fluency and its key components significantly vary, which is best summed up by Freed's (2000) suggestion that fluency is in the ears of the beholder.

1.4.3 Pedagogical implications of research on fluency

Fluency is a key component of efficient language performance. It is a skill which involves the transfer of the speaker's knowledge into speech at an acceptable speed and without too many interruptions in the form of pauses and hesitation phenomena. It is a skill which typically characterises advanced speakers and is required of them at a high level at advanced language exams. Despite this traditional association of fluency with advancedness it ought to be seen as a continuum, as a developing process which ought to commence at the earliest stages of language instruction and be carried out all the way through to the advanced level. The rising numbers of highly competent speakers of English shows that fluency is not a pipe dream but a realistic, fully attainable goal. Can fluency be taught, and how can research findings inform teaching practice?

The research clearly indicates that fluency has a psychological, cognitive basis, that it is related to automaticity. The way automaticity is developed is well described in psychological literature and if we accept that fluency is a skill like any other (Dörnyei 2009) than it can be developed using the principles known from other fields. These are based on the models of proceduralisation of declarative knowledge and its automatization (Segalowitz 2000; 2010).

1.4.3.1 Fluency development from the early stages of learning — repetition, recycling, using and reusing

Nation (Nation 2001, 205) wisely suggests that beginners "should develop fluency with greetings, numbers, time, days of the week, time indicators like today, yesterday, next week, last month, some colours, and other items which could be used frequently". Frequent use involves

repetition[10] and it is essential that the teacher constantly recycle previously covered language (be it vocabulary or grammatical structures). He should do so firstly in his own speech, secondly by eliciting known language, and thirdly by providing opportunities for independent use of known language.

As to the first of these three points, it is advisable for the teacher to keep a logbook of class vocabulary for each of the classes he teaches. Students can share the task of recording newly introduced words and phrases and the teacher ought to familiarize himself before the lesson starts with the vocabulary which ought to be included. New language introduced in the class ought to remain visible on the whiteboard so that the teacher can return to it whenever possible and so that it can help the students in communicative activities.

As to the second point, the teacher should frequently ask such questions which require the recycling of previously learnt language. Frequency is the key here, but authenticity helps to ease the pressure to perform. By authenticity I mean exploiting natural situations in the classroom, e.g. asking simple questions about the learners' world (e.g. *What day is it today? What is the time? What colour is my pen? What time did you go to bed yesterday?*). Authenticity is important in that it helps the learner focus on communication more than on the actual process of production, thus helping the learner speak without too much conscious awareness and attention. The teacher can ask two or three students the same question to provide an example, and then instantly instruct the learners to engage in a pair-work activity asking each other the same question. Such an activity, which is in accordance with Nation's (1999) principle of rehearsal and repetition, need not last long and can be introduced many times during each lesson. The language practised in this way should be familiar and simple so that the students do not need to pay too much attention.

As to the third point, the teacher must make sure that classroom activities do not only practise newly covered language but return to and incorporate what the students already know. It is

[10]Segalowitz (Segalowitz 2010, 212) warns that if repetition is to be successful "consistent association between words and meanings" has to be provided.

even helpful to return to exercises or texts which have already been completed. A dictation can be a useful technique of reintroducing older texts. Regular tests present another opportunity for recycling.

In a study aiming to provide the means of measuring the developmental stages of proceduralisation, Towell et al. (1996) attributed the increase in fluency of students after a stay abroad to the use of memorised sequences. This was evidenced by the increase in the mean length of runs. Other researchers also suggest the considerable importance of formulaic language (e.g. Pawley and Syder 1983; Lennon 2000; Wray 2000; Nation 2009; Schmitt 2010; Wood 2012). Work on formulaic language should also commence at the early stages bearing in mind the important principle of frequent repetition. A natural way of teaching chunks, which Raupach (1983) calls "islands of reliability", is by teaching not only individual words but also related phrases useful at the respective proficiency level. Thus, a beginner encountering the word *home* should be introduced to phrases such as *at home*, *she's at home, go home* and possibly link the word to previous knowledge by mentioning the now international word *homeless*). Lewis (1993; 2000) and Lewis and Gough (1997) in their Lexical Approach suggest a variety of techniques thereof. Teachers must ensure that the phrases are produced actively by the students and not just heard or recorded in their vocabulary books. Nation (2001) also recommends encouraging the students to perform faster when they are dealing with familiar and especially previously rehearsed tasks.

At word level automaticity can be compromised by pronunciation problems (Levis 2008), as speakers who need to concentrate on aspects of pronunciation deplete attentional resources which could otherwise be available for different production tasks. Pronunciation practice should constitute a regular component of lesson time and should include not only segmental but especially suprasegmental features. It is also useful to practise the pronunciation of whole chunks of language and not just single words. In this way we develop both *cognitive* and *performance fluency* (Segalowitz 2000).

Hilton (2008) and Milton (2009) stress the importance of vocabulary, and especially large vocabulary, for fluent speech. Milton sees fluency as a function of the ease and rapidity of access to vocabulary stores and claims that between 6,000 and 7,000 words are necessary for oral fluency and

for the attainment of the C2 level proficiency (p. 250). Even more vocabulary is essential for fluent writing. He claims that the teaching of compensation strategies ought to be reconsidered as they are "not an adequate substitute for knowing the vocabulary." (p. 240). Repetition is essential in vocabulary acquisition and should be encouraged even if it is just at the level of learning word lists.[11] "Evidence suggests this can be a very successful way to develop a sizable L2 lexicon." (p. 242). Milton also observes that research findings regarding the effect of repetition on vocabulary acquisition should have implications on the design of language teaching materials, which ought to pay closer attention to the way they recycle vocabulary.

Another important aspect of vocabulary knowledge which facilitates fluency is a good working knowledge of derivational morphology. This relies on appropriate declarative rules which have to be automatized if they are to contribute to fluent production. This is another area which can be incorporated into early instructional practice both actively in simple word-formation exercises and in noticing and awareness tasks.

Brumfit's (1984) appeal to distinguish between fluency- and accuracy-oriented classroom activities led to nothing more than the teachers' taking a more tolerant view of performance errors. The so-called fluency-oriented activities typical of communicative language teaching consist largely in not interrupting the students and letting them get the message across, which does not necessarily translate to fluency development. Communicative language teaching, in itself a reaction to audio-lingualism and its hated and inefficient drills, strongly opposed repetition as unnatural and unauthentic. The authenticity it promoted was, however, a great progress in that it emulated real communication and thus enabled transfer-appropriate learning (Segalowitz 2010) which is key to the

[11]Milton (2009) recommends bilingual lists with the L1 expression given first and the L2 equivalent being what the student has to produce. This is an imitation of the productive process similar to the recall procedure which takes place during speaking. The evidence he provides is based upon a quasi-experiment carried out by Stoddard*, who compared two groups of French learners of English, one of whom learned vocabulary receptively (i.e. from French into English), and the other one productively (i.e. from English into French). Subsequently the subjects took a productive vocabulary test, in which those who had learned productively achieved higher scores. (* Stoddard, G.D. (1929) An experiment in verbal learning. Journal of Educational Psychology 20, 452-457)

development of automaticity. At the same time, its opposition to repetition and thorough practice was counterproductive in that respect. Consequently, authors of textbooks and teacher training materials have been rather slow in implementing the principles of true fluency development. Rossiter et al. (2010) carried out a survey of fluency activities in a large number of ESL textbooks and teacher resource materials. Amongst the recommended activities, they found a prevalence of free-production tasks, which are only one way of encouraging fluency development, and very little space provided for the practice of formulaic sequences, for rehearsal and repetition. Another area of deficiency they discovered was the lack of consciousness-raising activities in both types of the investigated materials. This is clearly an evidence of the usual detachment of research and practice. This is especially unfortunate if we consider that fairly concrete methodologies have actually been described in academic literature (e.g. Gatbonton and Segalowitz 2005; Nation 2009; Rossiter et al. 2010). The following section describes some of them.

1.4.3.2 Activities for fluency development

As mentioned above, the principles of activities for fluency development are largely based on skill learning theory and its view of automatization. This has been succinctly expressed by DeKeyser[12] (2007, 107, quoted by Dörnyei 2009, 288) as follows:

"Automatization requires procedural knowledge. Proceduralisation requires declarative knowledge and slow deliberate practice. The acquisition of declarative knowledge of a kind that can be proceduralized requires the judicious use of rules and examples."

Dörnyei further points out that the application of this sequence resembles the very traditional and recently much-criticized PPP approach (present → practice → produce) but suggests a number of deliberate improvements. He calls the three stages *declarative input stage*, *controlled practice*, and *open-ended practice*. The first allows for the presentation of rules but strongly encourages

[12]DeKeyser, R. M. (2007) 'Skill acquisition theory' in B. VanPatten and J. Williams (eds.): Theories in Second Language Acquisition: An Introduction. Mahwah, NJ: Lawrence Erlbaum, 97–113

the involvement of the students in inductive tasks which enable noticing and awareness-building using a variety of focus on form techniques.

The controlled practice stage focuses on building procedural knowledge, whose development relies on repetition. Repetition involves drills and for these to be motivating they need to be carefully designed so that they are interesting and natural and involve not only structural but also communicative drills, drills in role-plays, games, songs etc. A variety of technological facilities (CALL, IWB, portable devices, internet) may be exploited to increase the level of enjoyment. Importantly, practice ought not to involve "higher levels of information (e.g. unfamiliar input, too varied content)" so that attention can be paid to the tasks at hand (cf. Nation's (2009) concept of *easy tasks*). Feedback is essential. Including variation once proceduralisation has taken place facilitates the transfer of the acquired skills to other contexts and tasks, which does not happen automatically.

The boundary between the second and third stages is very narrow and one of the problems of classroom learning is that learners proceduralize at different rates. They also do not have a sufficient amount of time for the processes to be completed in the classrooms. The focus of the last stage is on recycling and fine-tuning the performance using the target structures. Dörnyei warns that this is the most problematic stage of the sequence because we do not know how systematic and frequent recycling should be, and how to develop the range of "highly-specific task skills". Moreover teachers often succumb to the feeling that new material has to be introduced in every lesson (Nation 2009, 163), a tendency which I have frequently called a grammar race and which reduces the amount of time available for open-ended practice.

Nation (2009) specifies three criteria for fluency development activities. They are to be meaning-focused, linked to the learners' previous experience and performed at a somewhat higher rate than would normally be the case. A typical example is the "4/3/2 technique", which is a rehearsal-repetition activity in which the speaker tells the same story three times, on each repetition reducing the time he needs to retell the story. Nation suggests the timing of four, three and two

minutes which is, however, too long — speakers do not manage to keep track of the planning in such long stretches of time. A more appropriate timing appears to be 60, 45 and 30 seconds.[13]

Formulaic language can be practised using the disappearing text technique (Rossiter et al. 2010, 590), during which a text is displayed on the screen, read out loud by the teacher, then by the students and then each turn formulas are deleted from the screen and students have to read the text at speed trying to supply the deleted text. The gradual deletion continues until the whole text disappears and students have to reconstruct the whole length of it. The ideal initial length of such a text is about 60 words. A number of useful and highly practical techniques for general fluency development and formulaic language acquisition are suggested by Wood (2012, 195–203).

One of the most elaborate and research-informed methodologies for promoting formulaic language is Gatbonton and Segalowitz's (2005) ACCESS (**A**utomatization in **C**ommunicative **C**ontexts of **E**ssential **S**peech **S**egments) technique. It develops out of a belief that communicative language teaching fails to provide repetitive practice especially in the area of phrases. ACCESS activities are sequenced into three phases: Creative Automatization, Language Consolidation, and a Free Communication Phase. This is similar to the structure of automatization described above. The first phase contains presentation and concentrated practice and its aim is to "engage [learners] in a task or tasks in which functionally useful utterances are used and elicited naturally and repeatedly" (p. 329). This is achieved through such tasks as problem solving, role-plays, games, and simulations. The aim of the second phase is to "strengthen learner control of problematic utterances elicited and practised in Phase 1" through "fluency, accuracy and grammatical discovery tasks". In the last phase tasks such as problem solving, role-plays, and games are used to "engage [the learners] in a free communication activity or activities that deal with topics compatible with those of the Creative Automatization Phase." More simply, an ACCESS lesson starts with a very narrow communicative topic while opportunities are created for repetition (e.g. by using mingle activities). Secondly, the language produced is analysed and practised so that foundations for declarative rules are made and

[13]This is based on a personal experimentation with the technique.

consolidated. Finally, the topic is slightly broadened so that more contexts are created for communication, and further repetition of the same material is naturally ensured.

Wood (2012) designed a syllabus for a fluency course which he subsequently tested in a real classroom. Whilst he managed to record a significant increase in fluency and the use of formulaic language at the end of the course, as there was no control group it is impossible to make reliable conclusions. More classroom research is needed to test the effectiveness of such approaches.

1.4.3.3 Implications for teacher training

Derwing et al. (2009, 554) point out that direct fluency instruction might be especially important to students in non-English-speaking countries who naturally do "not have access to many opportunities to speak English outside of class". The question arises whether their teachers have been trained to provide such instruction. A brief survey of the latest editions of some of the most commonly used teacher-training manuals (Harmer 2007; Scrivener 2011; Ur 2012; Watkins 2005 and their older editions) shows that fluency is only mentioned in the context of providing feedback and teacher intervention during fluency activities. Fluency activities as such are however not defined, the manuals do not deal with the explanations of automaticity and general features of skill acquisition. Whether teacher training courses offer relevant information would require research well beyond the scope of the current thesis. But a brief survey of teacher's books of well-established and commonly used course books provides the same picture. Consequently, teachers have very little opportunity to learn about fluency development. Such a lack of connection between research and classroom practice ought to be addressed in teacher-training programmes by offering SLA courses which should provide both the essential theoretical framework and thorough discussions of concrete practical applications of the topics introduced. Besides, there is certainly plenty of scope for the production of activity books: some of the publications listed above (especially Wood 2012) provide a selection of useful ideas but these are not sufficiently concrete to be directly used in the classrooms. More context-related and task-specific ideas need to be developed and published. Not only with the aim of providing classroom ideas but also because of pointing teachers in the right direction so that they can develop

their own ideas, learn to adapt existing activities, and implement fluency development instruction into regular classroom work.

1.5 Accuracy

Of the CAF triad, accuracy is generally considered to be the easiest to imagine and define. Most definitions centre around the notion of norm and the extent of compliance with it. Some explicitly mention errors, e.g. Foster and Skehan's (1996) "freedom from error", Ellis's (2008) "the ability to avoid error in performance", and Housen and Kuiken's (2009) "the ability to produce error-free speech". Others are more cautious with the negative implication of errors, and focus instead on the norm, e.g. Hammerly's (1991) "the degree of deviancy from a particular norm", Skehan's (1996) "how well the TL is produced in relation to the rule system of the TL", Nation's (1999) "how closely learners' language resembles accepted standards", Pallotti's (2009) "the degree of conformity to certain norms", Housen et alia's (2012) "the extent to which an L2 learner's performance deviates from a norm (i.e. usually the native speaker)", or mention both norms and errors, e.g. Housen and Kuiken's (2009) "the ability to produce target-like and error-free language". Further two concepts need to be defined: error and norm.

Although researchers often lament how difficult it is to define an error (e.g. Gilquin and De Cock 2011, 142), existing definitions are fairly similar and straightforward, mostly based on the idea that errors are best seen as deviations from a norm. It is the application of such a definition that is hard, the process of determining what should and what should not be considered an error while analysing a particular sample of learner language. Here, the difficulty lies especially in choosing the appropriate norm. With English this is notoriously problematic as there are so many legitimate varieties and no official prescriptive set of norms. The situation is doubly complicated with spoken language. As heritage of the written-language bias in linguistics (Linell 2005), the majority of academic and pedagogical grammars are based on descriptions of written language. Whilst modern grammars (e.g. Biber et al. 1999; Carter and McCarthy 2006) include passages describing spoken varieties these

still occupy only a small part of the publications and are mostly based on comparisons with written language.

In general language, *norm* may be defined as "an authoritative standard; a model; a principle or standard of correctness that reflects people's expectation of behaviour, is binding upon the members of a group, and serves to regulate action and judgement"[14]. Such a definition explains why with regard to language use the norm is frequently associated with language as it is produced by native speakers. Leaving aside the difficulty of presenting a definition of a native speaker, this at least provides a useful point of reference. Thus Lennon (1991b) defines an error as *"a linguistic form, combination of forms, or utterances, which in the same context, and under similar conditions of production, would, in all likelihood, not be produced by the subjects' native speaker counterparts in Reading, namely the young adult educated native speakers of British English who formed the vast majority of the student body which functioned as the subjects' peer group and L2 linguistic community in Reading."* This extremely explicit definition, which Lennon used in a study comparing advanced-learner and native-speaker English, forefronts the fact that language norms are strongly contextually bound and defined not only by the speakers' nativeness, but also by their social class, education, age, and situational constraints. By saying "in all likelihood" Lennon allows for the occurrence of errors in native speakers' speech.

A norm is necessarily an instrument of comparison through which a sample can be set against a collective practice of members of a particular milieu. If, however, we are to compare like with the like we have to consider which milieu to choose. Looking at learner language we could select native speaker language just as well as the language of other learners. This is the basis of the comparative fallacy debate started by Bley-Vroman (1983), who claimed that it was a mistake to study "the systematic character of one language by comparing it to another" (p. 6), that "the learner's system is worthy of study in its own right" (p. 4), has its "own internal logic" (p. 15), and is not just "a degenerate form of the target system" (p. 4). Whilst looking away from the target language rules and focusing simply on the salient characteristics of the interlanguage per se facilitates the study of its

[14]The Longman Dictionary of Contemporary English, Harlow, Longman (1991)

structural features, it blatantly disregards the influence of input (Schwartz 1997) L2 learners are exposed to from the very start. In the language classroom, this takes form both as explicit rules and concrete and usually authentic examples of the native variety. However useful the abstraction from the TL perspective and a sole focus on the interlanguage structure might be for the development of SLA theories, it is useless for language pedagogy. The shared features of interlanguage might provide a descriptive framework but they cannot make a norm to be observed by other learners, especially as we know that in the early stages of L2A the L1 has a powerful influence on the developing interlanguage. Consequently speakers of different L1s would end up making different TL norms which is rather absurd. Thus it would appear logical that native-speaker norms are used for L2 learning and teaching. Without a native-speaker norm we could barely differentiate between various levels of proficiency and describe advancedness.

1.5.1 Error in the context of SLA — contrastive analysis, error analysis

When applied linguistics became established in the latter half of the 1940s, one of its main functions was to provide scientific foundations for language teaching and for the development of teaching materials. Fries (1945) claimed that "the most efficient materials are those that are based upon a scientific description of the language to be learned, carefully compared with a parallel description of the native language of the learner" (p. 9). Along with Robert Lado, he believed that by comparing L1 and L2 specific difficulties L2 learners were likely to have could be predicted. Predicting difficulties was an important step towards obviating errors which should be avoided at all cost during learning. The theory eventually came to be known as Contrastive Analysis Hypothesis (Wardhaugh 1970). Its key premise was that what played the most important role in L2A was the learner's mother tongue, which could affect acquisition both positively and negatively through processes of positive and negative transfer. Those features which were similar in the mother tongue would facilitate learning, those which were different would complicate it. The learner's L1 was thus seen as the primary source of L2 errors. It was soon discovered that difficulties were far harder to predict than originally thought and learners were found to be committing unpredictable errors and

not necessarily erring where expected. The research however marked the beginning of the scientific study of error and its causes.

The lack of success of contrastive analysis in explaining language acquisition from a behaviourist perspective inspired new theories which saw language learning not as imitation and habit formation but as an active and creative process during which rules are formed. Errors are no longer seen as degenerate forms but as instrumental in the process and as a natural and unproblematic developmental stage (S. P. Corder 1967). As such they became the focus of study motivated by the belief that the understanding of their nature would shed light on L2A. At the centre of this research stood a methodology called error analysis, which chose for its point of reference the TL norms. The key premise was that learner language had at its base an underlying rule-governed, dynamic system.

In his seminal article "The significance of learners' errors", Corder (1967) recognized that not all errors were to be treated in the same way. Some were only random one-offs, and the learner might even be aware of them and could correct them if asked. Others occurred repeatedly, they were systematic and unrecognized by the learner who could not correct them even if prompted. Corder called the first mistakes, and the latter errors.[15] Mistakes had little descriptive value for SLA whilst errors were indicative of "transitional competence", of what formed the "idiosyncratic dialect" (Corder 1971) of the learner. In reality, distinguishing between systematic and unsystematic errors is problematic.

---

[15]Cognitive psychology would classify Corder's mistakes as slips, as actions-not-as planned (Reason 1990), whilst Corder's errors are generally referred to as mistakes. In cognitive psychology they are defined as follows:

"Error will be taken as a generic term to encompass all those occasions in which a planned sequence of mental or physical activities fails to achieve its intended outcome, and when these failures cannot be attributed to the intervention of some chance agency."

"Slips and lapses are errors which result from some failure in the execution and/or storage of an action sequence, regardless of whether or not the plan which guided them was adequate to achieve its objective."

"Mistakes may be defined as deficiencies or failures in the judgemental and/or inferential processes involved in the selection of an objective or in the specification of the means to achieve it, irrespective of whether or not the actions directed by this decision-scheme run according to plan." (Reason 1990, 9)

Unlike contrastive analysis which compares native language and target language only, error analysis consists in comparing concrete samples of learner language with the target language. Corder recommends that error analysis should have six steps. The first one involves the collection of learner data.

The second stage consists in the identification of errors. These have to be defined, errors ought to be distinguished from mistakes, and overt errors from covert ones (overt errors are clear deviations in form and are detectable even out of context, whilst covert errors may be well-formed but do not actually communicate the speaker's intention). Owing to the fuzzy definition of error, this is the hardest stage during which complex decisions have to be made concerning the distinction between grammaticality, appropriacy and acceptability. Dušková (1969, 13) points out that there is a degree of deviation and acceptability, a continuum covering the span between fully correct, not incorrect but not entirely native-like, and fully erroneous (Lennon 1991b). Ellis and Barkhuizen (2005) speak of *absolute* errors, which are objectively erroneous, and *dispreferred* errors, whose acceptability depends on subjective judgement. Gilquin and De Cock (2011) suggest introducing yet another possible continuum — between error and dysfluency.

The third stage of error analysis involves the description of errors. This is based on their comparison with the reconstructed utterance in the target language and may be based on either linguistic or surface structure taxonomies (James 1998). Linguistic categories may distinguish between grammatical, lexical, phonological, syntactic and other errors, and these can be further subdivided. As a result different types of errors can be counted and their frequency evaluated. Lennon (1991a) shows that lexical errors are more frequent than grammatical (see also Ellis 2008, 50; Agustín Llach 2011, 70). Dušková (1969), on the contrary, finds errors in articles and morphology exceed those in lexis.

Surface structure taxonomies (Dulay et al. 1982) are behavioural taxonomies which focus on observable features of the erroneous behaviour: the formal characteristics of the utterances such as omission, addition, misformation, misordering, and the later added blends (James 1998). Such a classification is too mechanical and has little value for forming pedagogical implications.

Corder (1974) considers the systematicity of errors and suggests a distinction between pre-systematic errors (as a result of ignorance of the existence of a rule), systematic errors (a systematic application of the wrong rule), and post-systematic errors (an inconsistent use of a rule which has not been fully automatized). However much this taxonomy reveals the underlying principles of language acquisition in reality it is hard to find the dividing line between the different types.

The fourth stage of error analysis is the explanation of errors which is to be an attempt at identifying the reasons why they were made. Of Taylor's (1986) four-fold system of error sources which included psycholinguistic, sociolinguistic, epistemic or discourse sources, it was psycholinguistic sources which received most attention in both SLA and language pedagogy research. Following the tradition of contrastive analysis, the learner's L1 was seen as a primary factor causing interference errors (Richards 1973). Dušková (1969) has shown how the lack of articles in Czech affects the acquisition of the English article system by Czech learners. In their writing she found article errors to be the most frequent type but pointed out that not all of them could be explained as simple interference. Some of the errors were not a result of failure to use an article but rather instances of selection of the wrong one as a result of the learners' applying a rule incorrectly. Such errors are known as intralingual errors (Richards 1971) or developmental errors (Richards 1971; Schachter and Celce-Murcia 1977) but distinguishing between interlingual and intralingual/developmental may not be straightforward (cf. the difficulty in deciding whether the wrong use of the past tense instead of the present perfect by speakers of languages which do not have the present perfect tense is inter- or intralingual).

James (1998) links intralingual errors to universal learning strategies and identifies the following processes: *false analogy* (also known as over-generalization, which is a process during which a learner applies a rule in instances where the use of different rule is called for. E.g. *talk—talked → speak—speaked*); *misanalysis* (an erroneous hypothesis about a linguistic feature, e.g. when a learner thinks that any English word ending in *-s* is plural); *incomplete rule application* (also known as under-generalization; e.g. failure to produce correct word order in embedded questions); *exploiting redundancy* (failure to supply all markers even though they may not contribute to meaning; e.g. *He*

don't know.); *overlooking co-occurrence restrictions* (producing infelicitous combinations of words as in untypical collocations); *system-simplification* (replacing a number of distinct forms with a universal one, e.g. the use of *byself* as a universal reflexive pronoun). Whilst such a taxonomy can easily be illustrated with carefully selected examples the real problem arises when concrete samples of learner language have to be analysed as there is significant overlap between the categories.

The source of some errors may also be linked with the process of instruction. These are called induced errors and may be the result of faulty explanation (Stenson 1974) or inappropriate practice, such as overdrilling (Svartvik 1973). Whilst the identification of the source in this case is especially difficult as it would have to be traced to a particular classroom, data from learner corpora can reveal patterns of practice in presentation in textbooks and consequently in classrooms. Thus Granger (1999) suspects that many tense errors are teaching induced as "great many English grammars over-emphasise the role of adverbials in their presentation of tenses" (p. 197). Consequently learners develop strong associations between a particular tense and a certain adverbial which then acts as a trigger for the tense even when a different tense is required. Granger further points out that grammar instruction is guilty of presenting and practising aspects of grammar as isolated sentences, failing to show that tenses are part of cohesion. Such practice may eventually lead to errors in the use of tenses. Although teaching-induced errors do not reveal much about the process of acquisition they clearly have strong pedagogical implications.

Meara (1984) is sceptical of the value of error taxonomies owing to their low predictive and explanatory power, but as Agustín Llach (2011) points out such taxonomies offer preliminary data which facilitate subsequent analyses and help deduce behavioural and developmental patterns of errors. This can inform not only the design of classroom activities but especially of language textbooks.

The last stage of error analysis as suggested by Corder was to be error evaluation. Compared with the previous stages, it shifts the focus from the learner to the impact errors have on the addressee, evaluating the so-called error gravity. This is typically evaluated using external raters who use scalar judgements (e.g. Likert) to classify errors. Especially overall comprehensibility is considered.

Distinction is made between global and local errors. Local errors affect single elements in a sentence, whereas global errors go beyond them. Global errors are found to have a greater effect on comprehensibility. Another point of view considers acceptability which may involve judgements not only about well-formedness but also about contextual appropriateness. Besides these two criteria, Khalil (1985) considered the emotional effect on the addressee and suggested introducing judgements of the level of irritation errors may cause.

Many experiments have been carried out comparing the judgements of native and non-native speakers, and also those of expert (e.g. teachers) and non-expert judges. In a famous study, James (1977) compared the judgements of native- and non-native-speaker teachers and discovered that the latter were much harsher in their evaluations. Hughes and Lascaratou (1982) carried out a similar study but included also non-teachers and observed that non-native speakers appeared to be more concerned with accuracy whereas native speakers with intelligibility. McCretton and Rider (1993) claim that what conditions the teachers' assessment of errors is above all the syllabus used in their training or in their current practice (see also Davies 1983). The greater explicitness of syllabi used by non-native speakers would explain why they use more severe judgements.

Ellis (2008, 46) points out that the strictness in evaluating errors differs depending on whose performance is assessed. Whilst L2 learners' errors are seen as "unwanted forms", errors committed by native-speaker children are called "transitional forms" and those committed by adult native speakers are simply slips of the tongue. This shows that L2 learners' errors are more likely to be seen as errors of competence and native-speakers' errors as errors of performance.

The value of error gravity studies is questionable. In principle, qualified judgements ought to be based on clearly specified criteria. However, these have proved to be impossible to provide (see e.g. Johansson 1973). As a result only limited conclusions may be formulated in the form of guidance to teachers as to which errors they should pay greater attention to. This is unequivocally errors that impede communication which is somewhat of a foregone conclusion. Ellis and Barkhuizen (2005) nevertheless alert to the importance of teachers' developing their own understanding of the seriousness of errors so that they can be consistent in providing effective feedback.

1.5.2 A critical review of error analysis

As evidenced by Spillner's (1991) extensive bibliography of hundreds of error analysis publications, error analysis played an important role in the development of SLA theories. It has also received much critical attention especially owing to inherent methodological problems in failing to provide accurate and clear definitions of its key concepts, and in failing to ensure that all learner data analysed in the respective studies was homogeneous and truly comparable. As a result the studies are hard to replicate (Ellis 2008).

Another frequently mentioned point of criticism is the limited scope of error analyses which concentrate only on what the learners cannot do while not paying attention to what they are capable of. It fails to see which problematic areas the learners avoid, which features they underuse or overuse. It also ignores comprehension errors (Dušková 1969). Such criticism fails to see that the point of error analysis is not to present a complete picture of learner language but to learn to understand the principles of acquisition through exploring the areas where learner language fails to meet the parameters of the target language.

The limited scope of error analysis studies is best evidenced by the disproportion between studies exploring grammatical and lexical development. Dušková (1969) observes that lexical errors present "a much less homogeneous material for study than errors in grammar", which might be the reason why the majority of error analysis studies neglect lexical issues and focus rather on aspects of grammar (Agustín Llach 2011). Notable early exceptions are the studies by Meara (1984), Ringbom (1987), Lennon (1990a), Schmidt (1992), and Nation(1999). Agustín Llach (2011) provides a survey of studies from the 1990s and 2000s and shows that this is a fast developing area of research.

Another limitation is that error analysis offers a rather static picture of learner language (Dagneaux et al. 1998) with most studies being of cross-sectional nature. Štindlová (2011) points out that this is caused by the unavailability of longitudinal collections of learner language and is not an inherent fault of error analysis.

The main strength of error analysis is that it has changed the behaviourist view of errors as undesirable instances of language which are to be avoided at all cost to a mentalist understanding of errors as a necessary part of linguistic development and a possible evidence of the learners' built-in syllabus. In this way errors are seen from the perspective of language acquisition and processing, with the learner making "a significant contribution to learning" (Lennon 1991a). It is only when this perspective is adopted that error analysis can have an undeniable potential for language pedagogy. Modern approaches like computer-aided error analysis, which makes it possible to study much larger collections of errors using well-established corpus linguistics procedures, have enabled the production of error-aware teaching materials, such as the *Longman Dictionary of Common Errors* (Turton and Heaton 1996), *Learner English* (Swan and Smith 2001), and the *Macmillan English Dictionary for Advanced Learners* (Rundell 2007) with its 'Get-it-Right' usage sections which specifically target most frequent errors retrieved from learner corpora.

1.5.3 Measuring accuracy

The most widely used measure of accuracy is the percentage of error-free clauses (e.g. Foster and Skehan 1996; Skehan and Foster 1997; Mehnert 1998; Wolfe-Quintero et al. 1998; Yuan and Ellis 2003; Tavakoli and Skehan 2005; Ellis 2009). Whilst this is fairly unproblematic in written language, in spoken language it is much harder to establish what constitutes a clause. Even in written language of less advanced learners, identifying clauses may present problems. Consequently, other units of measure have been suggested, for example, the T-unit (Hunt 1970) and c-unit (Loban 1966) for writing, and the AS-unit for speaking (Foster et al. 2000; Ferrari 2012). Frequencies of errors can thus be calculated per unit of measure as a general measure of accuracy (Bygate 2001). This measure, however, fails to take into account the number of errors occurring within one unit (a unit which is not error-free may contain one but also more than one error). The question arises how to define an error-free unit if we do not provide a sound definition of error. This, as Polio (1997) observes, is rarely done by authors of accuracy studies.

Error frequency can also be given as an error rate expressed as an average number of errors per hundred words (e.g. Mehnert 1998; Sangarun 2005; Guara-Tavares 2008; Brand and Götz 2011; Götz 2013). This measure obviates the difficulty with defining and demarcating units, and it is easily applicable to both written and spoken language providing that plausible definitions of words are given. No distinction is made between more or less severe errors.

Wigglesworth (1997) works on the assumption that the acquisition of certain morphological features (such as verb tenses or plurals) can predict overall accuracy. He therefore expresses accuracy as the percentage of target-like verbal morphology. Working on the same principle, Crookes (1989) calculates target-like use of plurals, Skehan and Foster (1997) and Wendel (1997) target-like use of vocabulary and Ortega (1999) target use of noun modifiers and articles. The validity of these measures has been questioned (Ellis and Barkhuizen 2005, 151) as there is no fixed order in which morphemes and lexemes are acquired. The measures can, however, be successfully used to evaluate specific tasks which are designed to elicit particular morphological or lexical features.

A questionable measure of accuracy with a low validity is Wigglesworth's (1997) ratio of self-corrections as these rather point to the degree of control the speaker exercises over the process of monitoring his speech production. The question rather arises whether self-corrected errors should be included in the overall counts. The same applies to the question whether every occurrence of error should be counted even if the error is repeated more than once. Studies often fail to mention these issues specifically.

Accuracy on its own is not a reliable measure of language development as even low-level students can produce accurate language but the range of contexts and tasks will be limited, as will their fluency and complexity. With increasing proficiency, however, and especially at the advanced level, the descriptive role of accuracy increases (Norris and Ortega 2003, 737). For both low- and high-level learners, accuracy is an important indicator of the degree of control the learner has over the resources he has acquired.

1.5.4 Errors and language pedagogy

Errors have always played a key role in language pedagogy and related research and their view has undergone many changes. While it would be well beyond the scope of this thesis to provide a complete overview of attitudes to errors in the history of language teaching, a brief look at some of the most important ones will help understand the current position of accuracy in language teaching.

The troubled view of errors in language learning can be traced back to Quintilian (35–96 A.D.) who spoke for early exposure of children to Greek and Latin and maintained that good models of language use were essential as he thought errors would be difficult to erase later. Similarly, Comenius wrote *"The first attempt at imitation should be as accurate as possible, that not the smallest deviation from the model be made. That is to say, as far as is possible. For whatever comes first is, as it were, the foundation of that which follows. If the foundation be firm, a solid edifice can be constructed upon it, but if it be weak this is impossible. ... in any operation an error at the beginning vitiates all that follows ... Errors must be corrected by the master on the spot"* (Comenius 1657 (1907), 199–200)

As the aim of the grammar-translation method (GTM) was to teach classical languages, the development of productive competence was left aside. If errors were important at all, then it was for two primary reasons: 1) Language learning was seen as a form of mental exercise which was believed to be beneficial for intellectual development — neither could be expected to be especially tolerant of errors.; 2) The results of learning were tested using discrete-point tests where there was no room for errors. The GTM developed no theoretical foundations for its practices. It was primarily concerned with explicit teaching but not with implicit learning. Thus it never considered and developed any understanding of the transition from explicit into implicit knowledge.

Although the Reform Movement turned away from written language to speech, it put emphasis on practising accurate pronunciation from the initial stages of learning — a practice which was thought to be in the interest of intelligibility. Similarly, the Direct Method saw accuracy as central and advocated self-correction techniques and zero tolerance to errors as these would be impossible to eradicate if left uncorrected.

The key principle of Palmer's *Principles of Language-Study* was that of habit formation which was a process that relied on accuracy as erroneous behaviour could also be habitualised. This was the dominant view in language teaching for several decades, and affected especially the practice of the audio-lingual method. With its stimulus-response-reinforcement techniques of drilling and over-learning and its behaviourist underpinnings the audio-lingual method held the belief that any behaviour, including non-target-like, could turn into habits. The method is important in our discussion as it is the first modern language teaching method which is directly linked to SLA theory. Moreover it was a theory primarily based on description and prediction of errors, the contrastive analysis (see section 1.5.1). The span of its popularity coincided with the work of contrastive analysts such as Lado, and whilst it was intolerant of learner errors it was based on the idea of frequency and entrenchment which play a key role in the acquisition of skills.

As pointed out in section 1.5.1 it was the work of Corder and the domain of error analysis that made a revolution in the way errors were seen not only in SLA research but also and especially in language classrooms. Corder moved from detailed consideration and description of errors of his initial studies to suggesting that the term *error* be abandoned altogether (Corder 1981). He comes to see interlanguage as a legitimate dialect which is not to be compared with native-speaker norms. The learners' idiosyncratic sentences are not to be considered ungrammatical and although "it is true that they cannot be accounted for by the rules of the target dialect, they are in fact grammatical in terms of the learner's language" (p. 19). Such an approach leads to a "shift of emphasis in teaching away from a preoccupation with the grammar of the target language towards a concern with communication in the target language." (p.78) Corder's views are in line with the rise of communicative and learner-centred methods of the late 1970s and 1980s.

Just as the audio-lingual method moved away from the realm of explicit knowledge towards that of automatization and implicit knowledge, the communicative language teaching (CLT) approach moved away from rote-learning and drills towards the construction of implicit knowledge through the use of meaning-focussed communicative activities. There is a strange similarity between the grammar-translation method and the CLT: whereas the GTM had no theoretical underpinnings

whatsoever, the CLT was firmly backed up by various theoretical frameworks such as systemic functional grammar, the speech act theory and the model of communicative competence and proficiency. Both approaches, however, lacked psychological foundations and unlike the audio-lingual method, which was linked to behaviourism, they were not based on any psychological theory of learning.

As regards errors, the CLT embraced the concurrent applied linguistics views of errors as developmental stones and grew to emphasize meaning over form. What the CLT theoreticians certainly did not have in mind was the rejection of grammar. Widdowson (1978, 15) claims that the CLT "combines situational presentation with structural practice". Similarly, Littlewood (1981, 1) insists that it "pays systematic attention to functional as well as structural aspects of language, combining these into a more fully communicative view". However, the classroom practice with its emphasis on authentic input and implicit learning strongly deemphasized focus on grammar, a point for which the CLT eventually came under fire.

Another factor which influenced the CLT's treatment of errors was the increasing emphasis put on fluency development. Even though Brumfit (1984) advocated a disciplined and consistent separation of fluency and accuracy practice this was not to detract from the role of accuracy. Yet, as we have pointed above, classroom reality was somewhat different and sometimes "anything goes" was the rule of the day. Hammerly (1991), noticing the relaxed attitudes to errors and the commonly appearing call for the freedom of students' "creativity" pressed for caution:

"Some people encourage SL students to be 'creative' and to engage in what I have called linguistic adventurism. But if one is 'creative' with a complex tool one doesn't control, and one's misuse of the tool is not promptly corrected, one will develop poor tool-handling habits. The SL classroom does not and cannot offer students the kind and amount of feedback that allows NL or even many SL acquirers in the field to be linguistically creative while becoming linguistically accurate. In the classroom, if linguistic accuracy is not part of all activities from the start, and if linguistic creativity is not restricted to ***the creative use of what the students know of the language****, not much accuracy will develop."* (p. 86, my bold)

His voice was one of many calling for the vindication of accuracy- and grammar-focussed instruction in reaction to the "unprincipled CLT" (Dörnyei 2009). One of the alternatives proposed was task-based instruction (TBI). However, as Ellis (2003) points out, the TBI is a but a branch of the CLT. I agree with Dörnyei (2009, 278) in that tasks make SLA research more focussed and its results more easily interpretable, but TBI provides no novel approaches to accuracy development.

If contemporary language teaching was to learn from history it had to take on board two key findings: 1) the so-called focus on forms consisting in the explicit teaching of rules without their application to meaning and communication does not lead to the development of communicative competence; 2) the meaning-based communicative approaches (focus on meaning) which rely on implicit learning do not foster effective development of linguistic and communicative competences. These findings have informed the approach called focus on form (Nassaji and Fotos 2011).

Focus on form (FoF) starts with what the learners can do in communication and then encourages principled work with the requisite linguistic matter (be it grammar, lexis, phonology, discourse, pragmatics etc.). Attention is thus paid both to situational meaning and linguistic form both of which contribute to accuracy. Ellis (2008, 879–880) provides a taxonomy of key FoF teaching options. Nassaji and Fotos (2011) illustrate the theoretical foundations of the FoF approach with a number of practical examples of classroom techniques. DeKeyser (2007, 12) alerts to the challenge of designing a FoF syllabus without reverting to a structural syllabus that merely teaches "the structure of the day" or "becoming obsessed with a mere focus on forms instead of focus on form."

It is early days for passing judgement on FoF. Whilst there is a large amount of research and literature documenting its effectiveness, we have yet to see how successfully it will be adopted by language teachers and especially teacher trainers. However, its emergence is certainly a clear sign of the fact that in language teaching yet again form has moved to the forefront right next to meaning as its fundamental component and prerequisite. The only notable exception perhaps is the — in this light — anachronistic English as a Lingua Franca movement, which seems to have frozen in the interlanguage and learner creativity debates of the 1970s, ignoring the fact that the questions of language teaching are general and do not only concern the English language.

1.5.4.1 Errors and the advanced learner

As teachers of advanced learners are very well aware, even advanced learners with many years of experience using English in and outside of classroom regularly make mistakes. Significant amount of research backs up their experience (e.g. Dušková 1969; Lennon 1991a; Lennon 1996; Granger 1999; Thewissen 2013; Götz 2015). Are their errors instances of backsliding (Selinker 1972; R. Ellis 1985), lack of automaticity (Dušková 1969), having reached a plateau (Flynn and O'Neil 1988), being unable to attain full mastery (Han 2004), or are they what some researchers call persistent (Osborne 2007), stabilized (Schumann 1978), ingrained (Valette 1991), systematic (Corder 1967) or fossilized (Selinker 1972, Han 2004)? Or are they resistant rather than persistent (Han 2004)? In the following section I will explore a small selection of some of the explanations I consider the most important.

One of the most influential theories explaining L2 erroneous behaviour is the interlanguage theory. Its foundation stones were laid by Corder (1967; 1971) and extended by Selinker (1972) who suggests that a "latent psychological structure" exists in the learners' brain and is activated by the attempt to learn an L2.[16] Five central processes are supposed to contribute to the process of rule acquisition: language transfer (see above), transfer of training, strategies of second-language learning, strategies of second-language communication, and overgeneralization of target language linguistic material. Any of these processes can singly or in combination result in fossilization, a situation in which a learner retains an erroneous form as an adequate replacement of the correct one. Fossilization is defined by Selinker (1972, 229) as a mechanism affecting "linguistic items, rules, and subsystems which speakers will tend to keep in their interlanguage productive performance, no matter what the age of the learner or the amount of instruction he receives in the target language." Such a definition

[16]The interlanguage is an intermediary dynamic language system lying between the learner's L1 and the target language. As the prefix inter- implies, the interlanguage is an incomplete version of the target language. Bialystok and Sharwood Smith (1985, 106) claim that the difference between the two is in the representations of linguistic structures and the procedures for accessing the knowledge. Erroneous utterances thus result from deficient knowledge, deficient control or the combination of both (Hamilton 2001).

strongly implies that fossilization affects only errors thus equalling "error fossilisation" (Lennon 1991a, 130). Hamilton (2001, 75), however, disputes such a narrow view claiming that if the theory is correct there is no reason why it should only apply to errors and exclude beneficial fossilization. Consequently, we could not use psychological explanations to account for the difference between erroneous and correct utterances. Whilst this appears logically sound, it ignores the fact that fossilization is a term developed to describe the lack of learning rather than the success of it precisely in situations in which further development has ceased despite conditions which should promote it.

Han (2004, 25–26) makes a comprehensive list of behavioural reflexes and causal variables which contribute to fossilization including absence or insufficiency of instruction; absence of corrective feedback; satisfaction of communicative needs; lack of or low quality of input; age and maturational constraints; L1 influence; lack of attention; inappropriate learning strategy; language complexity; lack of opportunity to use the target language; will to maintain identity; false automatization; reluctance to take the risk of restructuring; lack of talent; possession of a mature cognitive system; processing constraints; natural tendency to focus on content, not on form; avoidance; failure to detect errors; transfer of training; lack of verbal analytical skills; lack of sensitivity to input; socio-psychological barriers; and multiple factors acting in tandem. All of these may play a role in reaching and sustaining the ultimate attainment to a different degree, depending on other learner variables.

Han singles out two factors as most influential. They are maturational constraints and native language interference. It is only the latter that can be properly taken on board in language instruction. Learner corpora containing samples of learner language with the same L1 are particularly useful here as they allow the identification of idiosyncratic features characteristic of the particular language group. These can then be made the focus of attention in the classrooms. However, this might have the opposite effect. One of the possible problems contributing to the fossilization of these items might consist in the non-native-speaker teachers' experiencing the same difficulties as their learners as they themselves are but advanced learners of the same language. Thus they might be unable to spot

and help eradicate persistent errors. Combined with the fact that contemporary textbooks do not address specific language transfer issues, little space is left for accuracy development in these areas.

The fundamental problem with fossilization is that it is virtually impossible to prove empirically as we would have to record repeated occurrences of problematic features over a period of time which are resistant to external influences such as prolonged exposure or instruction. From a methodological point of view, this is no mean task. According to Long (2003), it is more accurate to speak about stabilization which could be described as temporary (i.e. not permanent as is implied by the term fossilization) cessation of L2 development. Such a view allows both for further progress to be made, and for potential attrition which learners experience when they are short of practice and/or exposure. It has also been pointed out that research has so far been unsuccessful in explaining why only certain structures fossilize and why only some learners are affected (Osborne 2007).

Fossilization is not the only factor affecting accuracy. Research in psycholinguistics and psychology and backed up by neurobiological evidence (Dörnyei 2009) points to the importance of explicit learning and its co-operation with implicit learning. The "uneasy relationship" between the two is, according to Dörnyei, the "core dilemma of instructed SLA". Krashen (1981) famously distinguished between acquisition and learning, stating that what has been explicitly learned cannot be acquired. Reality, of course, proves the opposite with people regularly internalizing knowledge which they were first introduced to through explicit instruction. DeKeyser (1997) suggests that explicit precedes implicit, and that declarative knowledge becomes automatic through the processes of conceptualization and proceduralisation. Explicit knowledge can be drawn upon when problems arise. N. Ellis (2005) illustrates this on the example of walking, which is an automatic process until something unexpected — like stepping on unsteady ground — happens. At such a moment we draw from the pool of explicit knowledge, evaluate the situation and respond. Such processes seem to be at play in the production of speech: when at the enormous speed of production the learner's speech outruns the planning resources and he runs into difficulty, explicit knowledge is called upon. The greater and better the store of explicit knowledge, the more resources there might be for solving

problems and maintaining accuracy at the same time. It is worth noting how inextricably close fluency and accuracy are in this process.

Dörnyei (2009, 171–4) explains how explicit learning supports implicit acquisition claiming that "explicit registration of linguistic information allows implicit fine-tuning" (i.e. explicitly noticing[17] a feature is a trigger for its acquisition); that "explicit practice creates implicit learning opportunities" (i.e. conscious practice results in implicit learning); that "explicit knowledge channels implicit learning" (i.e. explicit knowledge creates mechanisms for further learning); that "explicit rote learning can provide material for implicit processing" (i.e. memorization builds up implicit knowledge); that "explicit knowledge fills the gaps in implicit knowledge" (for example see the walking parallel above); and, finally, that "explicit learning increases the overall level of accuracy in implicit knowledge". The ability to produce language is seen here as a complex skill whose successful acquisition depends on the interplay of explicit and implicit knowledge. In this light, errors made by advanced learners could be explained by either insufficiency in the store of explicit knowledge or by an uncompleted transfer of explicit into implicit knowledge.

1.5.4.2 Pedagogical implications for accuracy development in advanced learners

Accuracy studies show that even amongst advanced learners there is a degree of variance. Some learners are more accurate than others and further progress can still be made. The key factors to consider when negotiating the question of advanced-learner accuracy development are the role of explicit knowledge, explicit instruction and feedback, the degree of the learners' active involvement, the nature of their errors, and the role of fossilization and other causes of errors.

As we have shown above, explicit knowledge is seen to be of paramount importance. In a skill-acquisition theory (see section 1.4.1 in the discussion of fluency), it is the starting point for skill development (Towell 2012). Advanced learners, however, tend to rely on implicit learning which in itself is not sufficient as language acquisition does not simply take care of itself (Lightbown and Spada

[17]cf. Schmidt's (2001) Noticing Hypothesis

2013, 195) but ought to be based on comprehension, production and active study. Explicitness thus involves both focussed and guided attention to various language features in the form of explanations, analyses and explicit corrective feedback where necessary (Towell 2012). The approach that best suits this type of instruction is Focus on Form (see section 1.5.4) as it enables to combine a broader communicative framework with detailed linguistic study and practice.

Schmidt (2001) makes a truly valid point when he suggests that the learner's role has to be particularly active so that attention is paid even to those language features which are "infrequent, non-salient, and communicatively redundant". This is especially true for advanced features whose frequency might be particularly low.

Explicit instruction in advanced-language learning seems to be especially important and effective (Lightbown and Spada 2013) as it is experienced teachers who can make qualified judgements about the selection of features to focus upon. These ought to be especially common errors, L1-specific errors,[18] complex language features, and low frequency features. Teachers ought to provide explicit corrective feedback and encourage the students to analyse their own performance in order to notice how it differs from the target-language use. Last but not least teachers have to provide ample opportunities for practice as this is a prerequisite for proceduralisation. As Hammerly (1991, 21) points out, practice of correct forms does not only make perfect but also permanent.

1.6 Summary

Fluency and accuracy might be difficult to define and measure but they constitute two dimensions which reflect better than any other the extent to which the L2 and the application of its rules have become automatized. Thus they are especially important for advanced-learner-language research. In the CAF model they offer opportunities to measure language performance using linguistic means and thus assess the extent of the underlying competence.

[18]Lightbown and Spada (2013) observe that this is especially important in mono-lingual classes where the same errors and non-target-like features are repeated and learners are less likely to notice that they are deviant.

2. The data — LINDSEI_CZ

The reference data for this thesis was collected as a contribution to the Louvain International Database of Spoken English Interlanguage (LINDSEI), a large-scale multi-national corpus of advanced spoken learner English. LINDSEI is the spoken counterpart to the International Corpus of Learner English (ICLE), which was initiated in 1990 under the auspices of Centre for English Corpus Linguistics at the Université catholique de Louvain. ICLE is a learner corpus of written English comprising of argumentative essays written by higher-intermediate to advanced learners (3rd- or 4th-year university students of English). With its 3.7 million words to date, it is currently the largest multi-national learner corpus, containing essays by students from 16 different countries. In 1995, work was commenced on its spoken counterpart LINDSEI. To date, 14 countries have completed their contributions, giving LINDSEI 1.4 million words of learner language, and work is currently being carried out by a further 6 universities.

Table 2–1 LINDSEI — national subcorpora

Completed subcorpora	Bulgarian, Chinese, Czech, Dutch, French, German, Greek, Italian, Japanese, Polish, Spanish (2 universities), Swedish, Taiwanese, Turkish
Work in progress	Arabic (Saudi Arabia), Basque, Brazilian Portuguese, Finnish, Lithuanian, Norwegian

LINDSEI is complemented by the Louvain Corpus of Native English Conversation (LOCNEC), which is a spoken corpus of 50 native speakers performing identical tasks as those in LINDSEI. Comparison is thus possible not only between different L1s but also with native production. In the current thesis, LOCNEC provided data for comparing selected features of native and non-native fluency. The results can be compared with a similar study by Götz (2013).

Each LINDSEI national subcorpus contains the minimum of 50 interviews, which are made up of three tasks. The total duration of each interview is approximately 15 minutes. The interview is preceded by a brief period of time during which the interviewees prepare for Task 1. In Task 1, the students speak on a chosen topic (see Appendix 3 for a detailed description of tasks). Task 2 is a dialogue covering common topics such as the student's interests, study experiences, career plans and

aspirations. In Task 3, the student is given a set of four pictures from which he has to reconstruct a story without being given time for preparation. Task 1 might be considered a planned task, Task 2 is dialogic and Task 3 is spontaneous. The division into tasks makes the data suitable for analyses of the effect of task complexity on L2 performance (Ellis 2005; 2009). The predetermined structure of the interview guarantees homogeneity of the learner data.

2.1 LINDSEI_CZ participants

LINDSEI is a corpus of advanced learner English. It uses an institutional definition of advancedness. Such definitions (Ortega and Byrnes 2008, 9; Thomas 2006, 105) work with the premise that the subjects in question have had to comply with certain institutional criteria which serve as a guarantee of their proficiency. In the case of LINDSEI, the interviewees are 3rd- or 4th-year students of English philology. For the Czech subcorpus (henceforth LINDSEI_CZ), 50 students (with Czech as their L1) in their 3rd or 4th year of study of English philology at the Faculty of Arts, Charles University in Prague were chosen. A breakdown of learner metadata may be found in Appendix 2.

Table 2–2 LINDSEI_CZ — participant metadata

	Mean age	Mean length of studying English at school prior to university	Mean length of studying English at university	Length of stay in an English-speaking country	L3
n = 50	22.5 years (SD = 1.6)	9.9 years (SD = 2.6)	3.4 years (SD = .9)	mean = 9.9 months (SD = 25.8) med. = 1.2 months	German (25x), French (14x), Spanish (7x), Dutch (1x), Italian (1x), Russian (1x), none (1x)
Female = 43					
Male = 7					

2.2 Data collection

The majority of the 50 interviews were recorded in autumn 2012 and 2013 in the recording studio of the Institute of Phonetics (Faculty of Arts, Charles University, Prague), only about 15% of the interviews were recorded using a dictaphone. All of the interviewees met the required criteria for institutional belonging. They all filled in a learner profile form, and signed a permission for the anonymised data to be used for research purposes. 50% of the interviews were recorded by a native English speaker, the other 50% by a Czech advanced speaker of English. In some cases the learners

reported a high level of anxiety which may have adversely affected their performance. A particular problem occurred with reticent students with whom the interlocutors had to take a much more active role than with others. Consequently, with some of the interviews the boundary between Tasks 1 and 2 may be somewhat hazy. Similarly, less talkative students produced less language in Task 3, making the interlocutor ask more questions to elicit richer response. Tables 2–3, 2–4 and Fig. 2–5 show the proportions of the individual tasks.

Table 2–3 LINDSEI_CZ — description of the data

Choice of topic for Task 1	Length of A & B turns[19] in tokens	Length of B turns only in tokens	Duration of A & B turns (hh:mm:ss)	Duration of B turns (hh:mm:ss)	Mean length of interview in tokens	Mean duration of interview (mm:ss)
Country = 22 Film/play = 18 Experience = 10	123,761	95,904 mean = 1,918 (SD = 407)	12:52:25	10:37:42	2,475 (SD = 386)	15:27 (SD = 2:14)

Table 2–4 LINDSEI_CZ — task proportions (B turns only)

	Task 1	Task 2	Task 3
Word count (B turns only)	40,584	42,850	12,535
Mean word count	812 (SD = 329)	857 (SD =2 84)	251 (SD = 85)
Duration	4 hours 26 minutes	4 hours 38 minutes	1 hour 32 minutes
Mean duration (mm:ss)	5:19 (SD = 1:53)	5:30 (SD = 1:43)	1:51 (SD = 0:39)

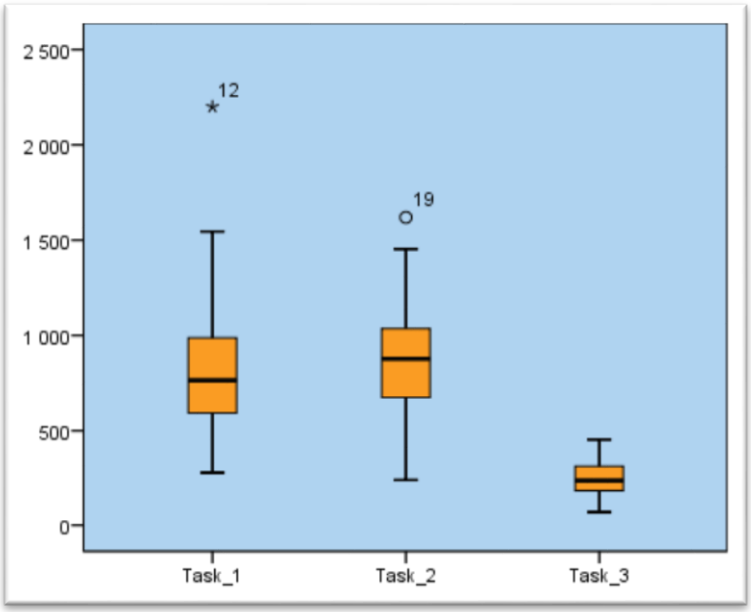

Fig.2–1 Boxplots showing the comparison of task lengths in words

[19] The phrase "A turn" and "B turn" denote utterances made by the interviewer and the interviewee respectively.

2.3 The transcription process

The transcription evolved in four stages. The first draft was the result of the speakers' transcribing their own interviews. This was subsequently checked by independent checkers, and finally the transcription was reviewed twice by the author of this study, who initially concentrated on verifying its accuracy with regard to the transcribed content, and in the last stage concentrated on the consistency of transcribing pauses and discourse markers. A set of computer scripts was designed for the purpose of a semi-automatic inspection of the consistency of all of the codes (e.g. the pairing of open and closing tags, and of the overlap tags, the correct format of all tags etc.), and the discovered errors were manually corrected.

The representation of spoken data in an oral corpus presents a major challenge in corpus linguistics research. Decisions have to be made especially as to how much non-verbal detail to transcribe, how to transcribe pauses, discourse markers, overlaps, prosodic features and other phenomena characteristic of spoken discourse. The transcribers of LINDSEI recordings are obliged to follow an official set of transcription guidelines (see Appendix 4), which uses both explanation, descriptions and exemplifications.

The transcription is orthographic. No punctuation marks are used. The transcription is not aligned but simply set out vertically with a paragraph corresponding to each turn. The speakers are marked as A (interviewer), and B (interviewee). A modified version of SGML[20] is used for presenting codes, which are placed in angled brackets: < > (opening tag) and </> (closing tag). The use of the closing tag is not required for the transcription of single phenomena such as sighs, giggles etc. Any data which would allow the identification of the speakers is anonymised. Empty pauses are transcribed as dots — short (< 1 second = one dot), medium (1–3 seconds = two dots), and long (> 3 seconds = three dots). Filled pauses and backchannelling markers are transcribed in round brackets distinguishing length and nasalisation as (eh), (er), (em), (erm), (mm), (uhu) and (mhm). Unclear passages are marked as <X> (one word), <XX> (two words) and <XXX> (three words and

[20]Standard Generalized Markup Language

more). Unclear word endings may be indicated by <?>. Truncated words are transcribed up to the point of interruption, which is marked using an equals sign. British spelling conventions are used, maintaining contracted forms as they appear in the recordings and some non-standard forms (*cos, yeah, kinda, gonna, wanna, gotta, dunno*). Foreign words are marked with tags (<foreign>, </foreign>). The only phonetic features transcribed are syllable lengthening (marked with a colon) and the stressed forms of indefinite (a[ei]) and definite (the[i:]) articles. Overlapping speech is indicated by the tag <overlap /> in both turns, however, spatial alignment and closing tags are not used. The following lines provide a small example of some of the features just listed.

<A> hello <first name of interviewee> </A>

<B> hello <first name of interviewer> <both laugh> </B>

<A> (erm) . are you nervous </A>

<B> . sorry what </B>

<A> are you nervous </A>

<B> nervous yes I I'm al= always always nervous <overlap /> when I have to talk </B>

<A> <overlap /> yeah always nervous really yeah </A>

2.3.1 Transcription limitations

Any transcription system is a compromise between validity and the ease of reading (Graddol et al. 1994, 185). Including too much detail obscures the text, including too little removes precious information. The level of granularity is determined not only by the type of research for which the transcription is intended, but also by its time and cost effectiveness.

The LINDSEI system of transcription strikes a good balance between the amount of detail and readability. Its limitations are especially the following:

– lack of alignment — this makes searching in the recordings extremely laborious;

– an unclear system of marking overlaps — this makes determining the duration of overlaps and the transcription of more complex overlaps difficult. As a result the transcriptions are not ideal for carrying out detailed conversation and discourse analyses.;

– insufficient definition of a pause — pauses are defined in the manual as "blanks on the tape" but no lower limit for the minimum length of pause is set. This presents a serious problem for pausology research.;

– lack of temporal information — this makes it difficult to measure the amount of language produced within a particular time limit and observe the variability of speech rate.

2.4 Error tagging LINDSEI_CZ

Computer-aided error analysis (CEA) relies on the retrievability of errors from a learner corpus. To this end, learner corpora are usually error-tagged. Error tagging is a complex process which presupposes a clear definition of language error and a system of error classification. As was shown in section 1.5, both areas are problematic especially in the area of spoken language. However, as CEA facilitates the study of learner language and its idiosyncrasies, error tagging is of paramount importance. Corder (1974) describes the process of error analysis as consisting of 5 steps: 1. collecting samples of learner language; 2. identification of errors; 3. description of errors; 4. explanation of errors; and 5. error evaluation. The process of error tagging as it is carried out in the current thesis represents steps 2 and 3. It identifies and describes errors but it does not attempt to explain them. The purpose of tagging is to provide searchable data for subsequent study of errors and infelicities which is to be carried out by linguists or teachers. During this process, some previously tagged errors may be rejected.

2.4.1 The Louvain error-tagging system

The tagging system used in this thesis is an adaptation of the Louvain error-tagging system (version 1.3) as described in Dagneaux et al. (2008), which was developed for the tagging of the ICLE corpus (see Dagneaux et al. 1998). Whilst it has certain limitations (it does not define errors and it was designed for assessing written language), it has been successfully used by other LINDSEI researchers (e.g. Granger and Thewissen 2005; Brand and Götz 2011; Kämmerer 2012; Götz 2013;

Thewissen 2013) whose results will thus be comparable with ours. The decision also strengthens the position of LINDSEI_CZ within the family of the existing Louvain learner corpora.

The Louvain system uses a descriptive, incremental annotation system based on linguistic categories.[21] It contains 55 error tags which are arranged in 8 categories:

– F (Form) — errors in spelling, and coining non-existent word forms;

– G (Grammar) — deviations from the rules of English grammar;

– X (Lexico-grammar) — errors stemming from ignoring morpho-syntactic properties of lexical items;

– L (Lexis) — errors in the selection of lexical items;

– W (Word redundant/missing & word order) — use of superfluous words, omission of words, ignoring the rules of word order;

– P (Punctuation) — missing, superfluous or wrongly used punctuation;

– S (Style) — use of unclear or incomplete sentences;

– Z (infelicities) — problems with register, political correctness, or stylistics.

These categories are encoded in the first position of each tag using the highlighted letters shown above. The second, third and fourth positions are occupied by letters denoting further subdivisions. For the category of Grammar, the second position of the tag provides information about the affected part of speech or word class (A=article, ADJ=adjective, ADV=adverb, D=determiner, N=noun, P=pronoun, V=verb, WC=word class). The tag GV thus denotes a grammatical error affecting the verb. The third position provides further information about the classification of the error, so within the GV group, for instance, AUX stands for auxiliary verbs, T for tense, V for voice etc. The tag GVT thus denotes a grammatical error consisting in the wrong use of the tense (see Fig. 2–2 for a graphic illustration of the principle of the tagging system). The fourth position is rarely used.

[21]The system does not encode explanation of errors (with the exception of the LSF tag, which denotes the use of a false friend) or attempts at grading them.

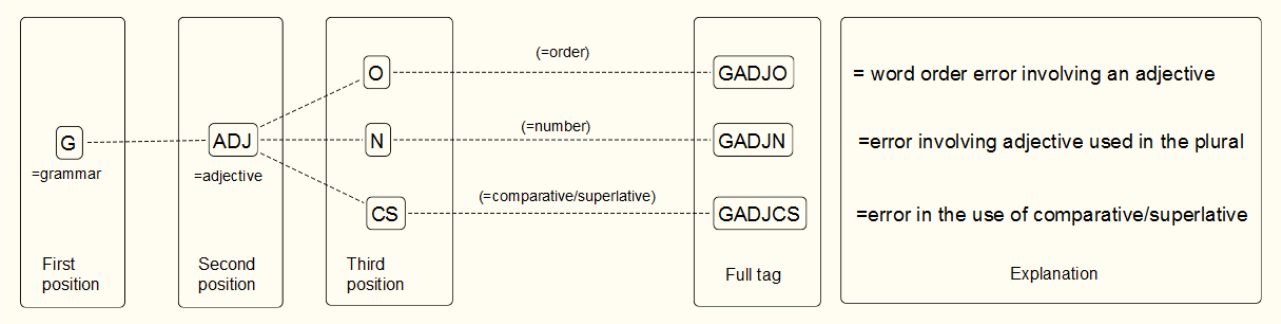

Fig. 2–2 An illustration of tag positions in the Louvain error-tagging system

The tags are inserted in round brackets immediately before the erroneous form.[22] Next to the erroneous form a suggestion of the correct form is inserted between the signs $. Only one corrected version is provided as these serve mainly to indicate the reason for the choice of tagging.[23]

they told us when we left —> they told us when we (GVT) left $were leaving$

The symbol 0 is used to denote an erroneous omission (e.g. of an article) or, in the corrected version, a suggested omission.

that's hard question —> that's (GA) 0 a hard question

in the society —> in (GA) the 0 society

The advantages of the Louvain system are especially in the following areas:

– simplicity (a relatively low number of tags is used);

– ease of use (tags are inserted as text into a word-processed text file);

– comprehensibility (the codes are easy to interpret and remember);

[22]The tags (X*CO) and (X*PR) are an exception as they are placed not immediately before the error but before the word which triggered the error. (e.g. *I'm (XADJPR) interested about $interested in$ music.)*

[23]E.g. CZ031 –"*I didn't have money*" – example of an error which can be corrected in different ways ("I didn't have the/enough money"). In this particular case either correction works without a significant change of the meaning. There might be other instances when there might actually be different readings depending on the meaning of the substitution.

– incrementality and openness (coders may expand their classification by adding further positions and/or tags);

– focus on consistency (the accompanying manual contains detailed explanations and exemplifications;

– ease of processing of the tagged files for purposes of data retrieval (this can be easily done using both standard word-processing and specialist corpus software such as WordSmith Tools, AntConc, MonoConc etc.).

Whilst in these aspects the whole system is in line with Granger's (2003, 467) description of effective error annotation system, it is not free from some limitations. These are especially:

– its focus only on written language;

– its lack of definition of error;

– the overlaps between some of the categories (e.g. grammar, lexico-grammar and lexis);

– the reduced legibility of the tagged text;

– the complicated addition of further codes (existing error tags may be in the way of tags for coding fluency features).

These limitations have to be addressed by the annotator. The system has to be adapted to spoken language (see below), errors have to be defined by the annotator to reflect the purpose of his/her research, overlaps between categories have to be carefully considered and explained. The reduced legibility and the difficulty in entering further codes without making the text even less legible are a tax for an otherwise simple and flexible system which allows data processing and data retrieval without the need for specialist software.

2.4.2 Error identification and tagging

Following the method of Brand and Götz (2011) who "count every major breach of a grammatical or lexical rule as an error" (p. 258) I included every instance of a deviation from the norm of contemporary spoken British English as it is described in contemporary academic grammars especially in Carter and McCarthy (2006) and Biber et al. (1999). No distinction was made between

errors, mistakes and slips[24] as it is notoriously difficult to distinguish them from each other. Repeated errors are all tagged as it is important to be able to retrieve all instances of erroneous use. Additionally, repetition of errors might be a sign of their systematicity or fossilisation. The tagging process can thus be described as strict and conservative.

Tagging was omitted when an error was instantly self-corrected.[25] Self-corrections, however, ought to be tagged in the future as they are an area worth exploring. Characteristic features of spoken grammar (as described in Carter and McCarthy (2006), Biber et al. (1999) and Miller and Weinert (2009)) were also taken into account and many features which would be considered erroneous in written text were ignored. These included especially repetitions, topicalization, tails, reformulations, abandoned or incomplete utterances, declaratives used as questions, certain peculiarities of word order (e.g. *This is what he meant, I think, isn't it?*). Also ignored were errors based on cultural knowledge (e.g. CZ033 — **The Midsummer's Night's Dream*). Section 2.4.3 below describes some of the complex decisions made during error identification and tagging.

The error tagging of LINDSEI_CZ happened in four stages. The first one involved a manual identification and marking of all errors, infelicities and queries following a definition described in the preceding paragraphs. Errors were highlighted in a printed version and queries were discussed with a native speaker.

During the second stage, errors were assigned appropriate tags (see Appendix 4 for a list of tags used) and these were entered in the text. To this end, I designed an error editor[26] (ErroGraph v.1.0) as a script in the WordPerfect macro language (see Fig. 2–3 for a screenshot). This editor enabled a fast, easy and accurate way of entering the tags and corrections into the text. The editor is highly

[24]CZ005 – "here (WO) this in $in this$ seminar" – judging by the otherwise high proficiency of the speaker, this word order error is just a slip. Nevertheless, all such instances are marked as erroneous.
[25]CZ025 – "I love they that accent < overlap / > they have (erm)" – an example of self-correction
[26]An error editor was designed in the CECL for the annotation of ICLE. This, however, failed to work on my computers owing to an unresolvable conflict in a dynamic link library.

flexible and tags can easily be added or adapted. Whilst the editor is not an essential tool for error tagging it helps to ensure consistency and somewhat speeds up the process of tag insertion.

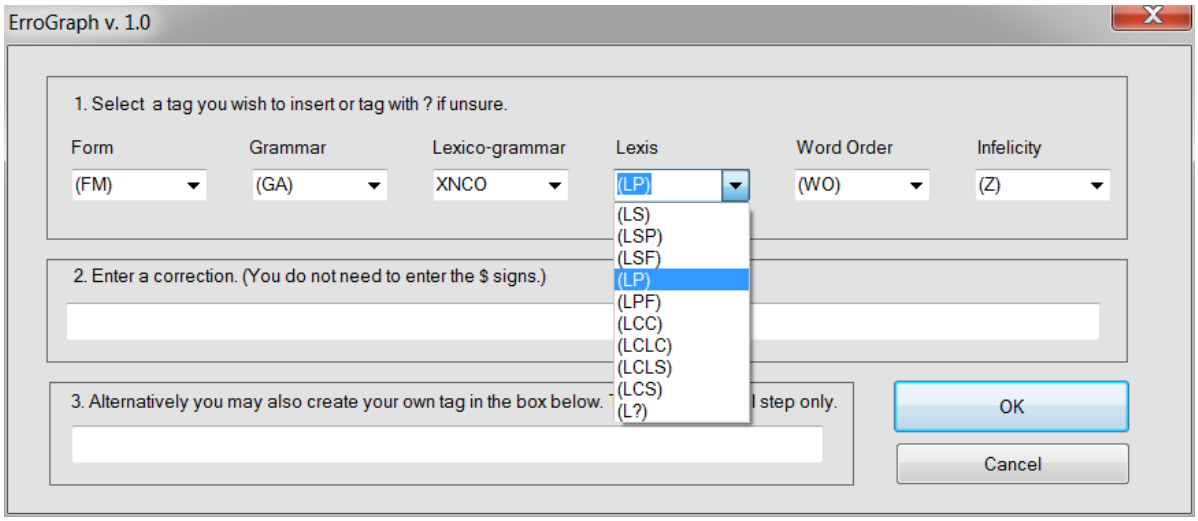

Fig. 2–3 A screenshot of the ErroGraph v. 1.0 error editor with a sample, pull-down menu

A query tag was introduced into the system to mark any features which needed a more thorough verification (see section 2.4.3 for a description of difficulties). A list was drawn up of features which appeared to require a different sort of tagging. This led to the addition of ten new tags (see Table 2–5). A complete list of tags used for tagging LINDSEI_CZ may be found in Appendix 5.

Table 2–5 A list of newly introduced tags in LINDSEI_CZ

New tag	**Explanation**	**Example**
GVAUXC	wrong auxiliary in the conditional	*if it (GVAUXC) would be $were$ better*
GVTC	wrong tense in a conditional clause	*if it (GVT) was $had been$*
GVMC	wrong form of the infinitive in the conditional	*I would (GVMC) do $have done$*
GVTA	tense agreement error	*he told us that (GVTA) it's $it was$ raining*
GADVI	intensifier error	*it's a (GADVI) very $wonderful$ city*
GNUM	an error involving a numeral or a numeric expression (but not a quantifier, i.e. the erroneous expression must contain a numeral)	*in her late (GNUM) thirteens $thirties$*
GEX	existential constructions	*(GEX) there was $it was$ dark there*
GPRO	wrong use of proform	*have you seen the film – yes (GPRO) I've seen $I have$*
LSP	wrong use of independent preposition	*it's (LSP) on in the second picture*
DTG	a double tag — this is used to mark cases of double errors which are difficult to annotate separately	*I realized that (DTG) (GVTA) (WO) actually I enjoy $I was actually enjoying$ it*

The majority of these new tags were added due to the seemingly high occurrence of these errors in my corpus. I was also guided by my experience of teaching advanced learners and predicting that these tags might help better target areas which I considered problematic. This applies to problems with the conditionals (GVAUXC, GVTC, GVMC), tense agreement (GVTA), intensification (GADVI), and independent prepositions (LSP). Problems with existential constructions, proforms and numerals were not expected to be numerous but appropriate tags seemed to be lacking in the Louvain manual. The (DTG) tag was adopted to agree with the error-tagged version of German LINDSEI (Kämmerer 2009). Other tags could be introduced to describe specific features of spoken language (see Gillová 2014) but here I concentrate only on instances relevant to error analysis.

During stage three, the whole of the corpus was annotated by a trained independent rater whose suggestions for changes were considered and where appropriate incorporated into the text. It was at this stage that queries marked with a query tag were resolved.

Stage four involved a final check of all tagged examples. Concordances for each tag were drawn up in the AntConc concordancer and all groups of examples sharing the same tag were thoroughly inspected. At this stage the error tagged version was considered complete and proper error analysis — quantitative and qualitative — could commence.

2.4.3 Difficulties encountered during error-tagging

Numerous difficulties had to be resolved during the error tagging process. Some of these stem from the fuzzy definition of error, others from the difficulty of identifying errors in spoken language. The following paragraphs illustrate some of these instances.

2.4.3.1 Errors allowing different tagging

CZ004 — *"say truth"* — I consider this as one error, as a wrong lexical phrase, i.e. (LP). It could also be tagged as two errors: (LS) and (GA). The (LP) reading is a better reflection of the fact that this is probably a transfer error, a literal translation of the Czech *"říct pravdu"*.

CZ005 — *"(GPP) me my switching law for English was a good decision"* — an example of colloquialism which is here classified as an error, but it could also be ignored as this usage is fairly typical in spoken grammar.

CZ024 — *"there is a (er) course English . (eh) what (GPR) how $what$ is it called"* & *"I started as (erm) (er) . (GPR) how $what$ is it called"* — an example of how difficult it is to categorise an error, as here it could be a (GPR) as I have marked it, but it could equally well be an (LP) as that is what is lacking in the student's repertoire (and she makes the same error twice) — if it is a (GPR) then we are looking at it purely grammatically, if it is (LP) then we look at the error from the perspective of what is missing in the student's repertoire, so classifying some error depends on the view that we take of errors.

CZ025 — *"just (eh) flat"* — another illustration of the difficulty of error tagging speech — is this a type of a nominal mistake, a mistake in the use of article (GA) or a false start, or is this a filled pause (eh)? On listening to the recording, the last option appeared to be more likely.

CZ026 — two instances of *"that was really fun"* instead of *"real fun"* — I tag it as an intensifier error (GADVI) but it could also be seen as an instance of a lexical error.

2.4.3.2 The need to check the recording

There were several instances when it was necessary to check the recording in order to make an accurate assessment.

CZ005 — *"I wasn't sure . not at all"* — whilst this appears in the transcription as a double negation error, the prosodic features make it clear that *"not at all"* is a separate non-erroneous clause here.

CZ024 — *"to be some kind of visually interesting"* — *a*n example characteristic of spoken language, a result of online planning, the intonation and the speech rate make it clear that the speaker was rephrasing. Such instances should not be marked as errors.

CZ030 — *"so most that was another problem most of the time I stayed in the house"* — checking the recording proves that this is an example of a false start which is resumed after the inserted clause.

CZ033 — *"I don't think I have a . really . passion"* — the pauses and intonation indicate that the speaker meant to use the disjunct *"really"* and not the adjective *"real"*. The idiosyncratic word order is not erroneous, the speaker interrupts her utterance, intensifies it, and then resumes it. This is a common feature of spoken language.

CZ035 — "so that's experience was also very surprising to me" — prosodic features make it clear that this is an example of online reformulation, a type of uncorrected false start. Such instances are not marked as errors.

CZ039 — *"the the studying literature opens your mind"* — here listening to the recording makes it clear that this is not an error with the definite article but a reformulation, a false start. Consequently, this is not classified as an error.

CZ047 — *"the audience makes a . amazing atmosphere"* — this is not an article error as the pause suggests that the speaker may have reformulated. It's possibly a sign of the adjective not being planned at the time of the utterance, otherwise the speaker would have used *"an"*. Consequently, I do not mark such instances as erroneous.

3. Methodology

The aim of the present study is to explore two dimensions of advanced L2 performance in English, namely accuracy and fluency, in order to identify characteristic features of advanced-learner performance especially as regards its deviances from the target-language use on the accuracy and fluency planes. Research findings indicate that even at the advanced level learners are short of the target and produce non-target-like features which may be stabilized or random but which allow space for further improvement. It is my aim to identify precisely what these are, how they correlate, and what their causes might be so that pedagogical conclusions may be formulated. What appears to be of particular importance is the discovery of any traits which are not just idiosyncratic but rather typical of the whole group or developmental stage.

The explorations will be based on a corpus-driven approach using both quantitative and qualitative computer-aided error analyses (CEA) and contrastive interlanguage analyses (CIA) (Granger 1998; 2015) whilst making use of native and non-native-speaker comparisons along with non-native-speaker comparisons per se. These comparisons are expected to reveal the strengths and weaknesses of the investigated learners' English, and also any typical patterns of their spoken language production. The afore-mentioned pedagogical conclusions are hoped to have effect both for the advanced level itself and for any of the preceding levels of proficiency thus affecting the whole learning process (cf. Cobb 2003).

The source of data best-suited to this purpose is a learner corpus — a "systematic collection of authentic, continuous and contextualized language use by L2 learners stored in electronic format" (Callies and Paquot 2015, 1).

The field of learner corpus research was established relatively recently, in the 1990s, as an off-shoot of general corpus linguistics. One of its founding figures, Sylviane Granger, spoke of its potential to address specific learner needs and foster the creation of specialized materials focusing on their difficulties (1994, 29). However, fifteen years later Granger (2009) points out that "learner corpus research has not yet fully realized its stated ambition as its links with SLA have been somewhat weak and it has given rise to relatively few concrete pedagogical applications" (p. 14).

Granger (ibid.) notes that this is due to the relative youth of the field where more time is necessary to perform analytic work and verify the findings. Nothing could be more true in the field of spoken corpora which are scarcer, smaller in size, but significantly more laborious to compile and analyse. This should, however, not detract from their potential for pedagogical applications.

The data source for my study is based on a spoken corpus of 50 advanced learners, students of English philology, and a parallel corpus of 49 native speakers. These relatively high numbers increase the chances that the findings will not be "applicable only to the one or two learners studied [... but will be] indeed characteristic of a wide range of subjects" (Mackey and Gass 2012, 9). The participants in both groups are of similar age and social background and perform identical tasks in similar settings. The corpora can thus be considered fully comparable (De Cock 2012).

The project evolved in several stages. Stage one consisted of making the recordings and transcribing them (see sections 2.2 and 2.3 for a detailed description). Stage two involved a thorough check of the transcriptions. The goal of stage three was to perform a small-scale pilot study in which a small selection of transcriptions were error-tagged, error rates were calculated and initial fluency measurements were made (the pilot study is described in section 3.1). During stage four the whole corpus was error-tagged (see section 2.4) and fluency components were measured for both the non-native and native component. The aim of stage five (described in Chapters 4 and 5) was to analyse the data using corpus linguistics tools (such as AntConc[27] and WordSmith[28]), audio tools (WavePad[29]) and statistical tools (SPSS). Statistical tests (e.g. parametric t-test, ANOVA) were used to provide descriptive statistics and the significance of the findings (these are reported upon in Chapters 4 and 5). The resulting tables and graphs were created in MS Excel and SPSS (see in Chapters 4 and 5). Stage

[27]Anthony, L. (2014) AntConc (Version 3.4.3) [Computer Software]. Tokyo, Japan: Waseda University. Available from http://www.laurenceanthony.net/

[28]Scott, M. (2014) WordSmith Tools (Version 6) [Computer Software]., Liverpool: Lexical Analysis Software. Available from http://www.lexically.net/wordsmith/

[29]WavePad Audio Editing Software (Version 5.96) [Computer Software], NCH Software, Available from http://www.nch.com.au/wavepad/index.html

six involved a comparison of native-speaker and learner fluency, and correlations were calculated using Pearson's r.

3.1 Pilot study

The pilot study was carried out partly as a feasibility study in order to determine the exact methodological steps for the full-fledged analysis, and partly to determine which concrete measures were going to be included. The pilot study has two parts: the first one deals with accuracy, the second one with fluency, research questions and hypotheses are introduced within the two sections.

At the time when the pilot study was carried out only 25 of the final 50 speakers had been recorded. A small selection of these was made based on an impressionistic, perceptive evaluation of the learners' proficiency which was acquired during the process of the audio checking of the transcriptions. The following six speakers and criteria for selection were used:

1. Speaker 1 (henceforth PS1) — the most proficient speaker (the most native-like accent, a fast speaker, very accurate and idiomatic language with virtually no trace of L1 transfer);

2. Speaker 2 (henceforth PS2) — the least proficient speaker (strong accent, many errors, slow speech, many self-corrections);

3. Speaker 3 (henceforth PS3) — the fastest speaker;

4. Speaker 4 (henceforth PS4) — the slowest speaker;

5. Speaker 5 (henceforth PS5) — the average speaker (not particularly outstanding performance but a sound, good command of the language with occasional errors, slight accent and good fluency).

6. Speaker 6 (henceforth PS6) — a below average speaker (stronger accent, less accurate)

3.1.1 Pilot study — accuracy

Accuracy was operationalised on the basis of error frequency (error rate calculated as errors per hundred words). Therefore, as the first step, the selected six transcriptions were checked for errors, which were classified according to the Louvain error-tagging manual (for detailed description

see section 2.4). A total of 214 errors were identified. This made it possible to evaluate the error rates for the individual speakers. The results are shown in Table 3–1.

Table 3–1 Pilot study — accuracy — error counts and error rates

Speaker	**PS1**	**PS2**	**PS3**	**PS4**	**PS5**	**PS6**	**Total**
Error count	11	88	23	15	33	44	**214**
% of total	5.14%	41.12%	10.75%	7.01%	15.42%	20.56%	**100%**
Word count	3045	2123	2126	1817	1990	2178	**13279**
Error rate	0.36	4.15	1.08	0.83	1.66	2.02	N/A

As the table reveals, PS1, the speaker who was perceptively selected as the most proficient has the lowest error rate of 0.36 errors per hundred words (i.e. committing one error in approximately 278 words). The weakest speaker, PS2, with an error rate of 4.15 errors phw, commits an error in approximately every 24 words. The speaker perceptively evaluated as the fastest, PS3, has an error rate of 1.08, committing one error in approximately every 93 words. The slowest speaker, PS4, with an error rate of 0.83 errors phw commits an error in approximately every 114 words. The average (PS5) and below average (PS6) speakers with error rates of 1.66 and 2.02 errors phw commit one error in approximately every 60 and 98 words respectively.

Accuracy measured in terms of error frequency provides a good picture of the accuracy spread in the selected sample but offers no information about the nature of errors. As the next step I therefore closely investigated the concrete errors and calculated the frequencies of the individual error types, i.e. groups F (formal), G (grammatical), X (lexico-grammatical), L (lexical), W (word order etc.) and Z (infelicities). I used the error concordancer AntConc to retrieve and count the individual error groups. The results are shown in Table 3–2 (overleaf).

Table 3–2 Pilot study — accuracy — error types

	F	G	X	L	W	Z	Total
PS1	0	6	1	1	3	0	**11**
PS2	0	40	2	32	9	5	**88**
PS3	0	12	1	7	0	3	**23**
PS4	0	17	2	12	2	0	**33**
PS5	1	27	2	11	2	1	**44**
PS6	0	8	0	7	0	0	**15**
Total	**1**	**110**	**8**	**70**	**16**	**9**	**214**
%	**.47%**	**51.40%**	**3.74%**	**32.71%**	**7.48%**	**4.21%**	100%

Grammar errors were found to be the most frequent, forming 51.4% of the total. Lexical errors were the second strongest group with 32.71%. These two groups far outweigh errors in the word-order group (7.48%), infelicites group (4.21%), lexico-grammatical group (3.74%) and the single formal error (.47%).

To obtain a deeper insight into the nature of the errors I then subclassified the errors within the G and L groups. As errors in the other groups were scarcer, I avoided subclassifying them in the pilot stage. Table 3–3 shows the classification and distribution of grammatical errors. For convenience sake several of the least frequent types of errors were grouped under the heading "Other".

Table 3–3 Pilot study — accuracy — classification of grammatical error subtypes (GA = articles, GVT(A) = tenses including tense agreements, GVM = verb morphology, GADVO = word order with adverbs, GDI = indefinite determiner, GVAUX = auxiliary/modal verb, GNN = noun number, GPP = personal pronoun)

	GA	GVT(A)	GVM	GVAUX	GADVO	GDI	GNN	GPP	Other	Total
PS1	2	1	1	0	0	1	0	1	0	**6**
PS2	15	3	8	4	1	1	1	0	7	**40**
PS3	7	3	0	0	1	0	1	0	0	**12**
PS4	4	4	0	0	0	0	0	0	0	**8**
PS5	9	2	0	0	1	1	0	2	2	**17**
PS6	12	1	1	2	2	2	2	1	4	**27**
Total	**49**	**14**	**10**	**6**	**5**	**5**	**4**	**4**	**13**	110
%	**44.55%**	**12.73%**	**9.09%**	**5.45%**	**4.55%**	**4.55%**	**3.64%**	**3.64%**	**11.82%**	100,00%

The six learners most frequently erred in the use of articles (44.55% of all errors) and tenses (12.73%). The other errors appear to be either more random or they are mostly committed by the weakest speaker PS2. A closer inspection of the article group revealed that of the total 49 article errors

26 (53%) instances involved an omission of the article. A closer analysis of the tense errors revealed that all 14 errors involved the use of the present perfect.

As regards lexical errors, the second most frequent type of errors, their classification and distribution are shown in Table 3–4.

Table 3–4 Pilot study — accuracy — classification of lexical error subtypes (LS = single words, LSP = prepositions, LP = phrases, LC = lexical connectors)

	LS	LSP	LP	LC	Total
PS1	0	1	0	0	1
PS2	8	8	14	2	32
PS3	2	3	1	1	7
PS4	2	5	0	0	7
PS5	1	7	4	0	12
PS6	3	0	8	0	11
Total	**16**	**24**	**27**	**3**	70
%	**22.86%**	**34.29%**	**38.57%**	**4.29%**	100%

As the table illustrates, the learners most commonly erred in the use of lexical phrases (38.75%), prepositions (34.29%) and single-word choice (22.86%). A deeper analysis of the LP and LS errors did not reveal any trends, unlike the LSP group, where 12 (50%) instances involved the use of the phrase "on the picture" instead of the target-like "in the picture".

Table 3–5 shows the effect of task design on accuracy. As task 3 is considerably shorter it would be misleading to use raw or relative frequencies. A more appropriate measure of accuracy seems to be a task error rate calculated as the total number of errors divided by the total number of words produced in the respective task by all of the pilot speakers and multiplied by a hundred. Task 1 has the highest error rate with 5.11 errors phw, Task 2 resulted in 3.85 errors phw, and Task 3 in 3.36 errors phw. Most (74%) of the errors in task 3 are lexical errors.

Table 3–5 Pilot study — accuracy — distribution of errors in tasks

Task	F*	G*	X*	L*	W*	Z	Total	%	Word count	Error rate per task
T1	0	67	1	25	10	4	**107**	**50%**	2095	**5.11phw**
T2	0	40	6	25	6	3	**80**	**37%**	2079	**3.85phw**
T3	1	3	1	20	0	2	**27**	**13%**	803	**3.36phw**
Total	**1**	**110**	**8**	**70**	**16**	**9**	**214**	100%	**4977**	**4.3phw**

3.1.2 Pilot study — accuracy — conclusion

The results show that the perceptive evaluation of the speakers' proficiency was accurate and that the range of error frequency is fairly wide, ranging between one error every 278 words and one error every 24 words. The error interpretation is somewhat skewed by the inclusion of speaker PS2 who makes 36% of all of the errors identified. Whilst including him in the count shows we might be able to expect a wider range of skills in the studied sample than the label advanced learner might imply, no general conclusions should be made at this stage about the characteristics of the pilot group's accuracy.

There appears to be a correlation (see Table 3–6 and Fig. 3–1) between the perceived proficiency (and measured accuracy as expressed by the error rate) and the number of errors in the use of articles. Also, lexical errors (and especially lexical phrase errors) seem to correlate with error rate. Errors in the use of tenses do not show this trend. The sample is, however, too small to evaluate these hypotheses using statistical tests.

Table 3–6 Pilot study — accuracy analysis — comparison of article (GA), tense (GVT), lexical (L), and lexical phrase (LP) errors with error rate*

	GA	GVT	L*	LP	Error rate
PS1	2	1	1	0	.36
PS4	4	4	7	0	.83
PS3	7	3	6	1	1.08
PS5	9	2	12	4	1.66
PS6	12	1	11	8	2.02
PS2	15	3	30	14	4.15

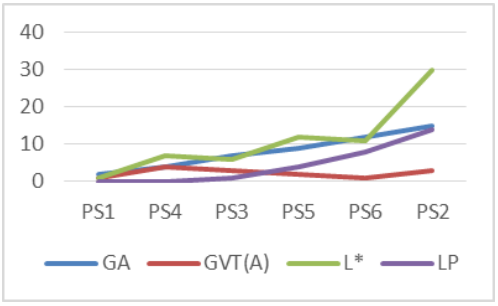

Fig. 3–1 — Pilot study — accuracy analysis — a graph showing the relationship between the increasing error rate and article (GA), tense (GVT), lexical (L), and lexical phrase (LP) errors*

It appears that accuracy is affected by the design of the tasks and thus it is essential to perform error analyses not only in the three tasks as a complete dataset, but also in the individual tasks.

The pilot study proved the feasibility of the selected method of measuring accuracy. Whilst errors for the purposes of spoken language analysis are notoriously difficult to define and identify with certainty, once they are found they can be evaluated using both quantitative and qualitative tests. Even a small sample such as the one selected for the pilot study provides information which can be further tested and empirically evaluated in more extensive studies.

The present study investigates advanced learner accuracy. Based on the result of the pilot study the following steps are suggested. First, the whole corpus will be error tagged. Accuracy will then be calculated for every speaker as an error rate expressed as the number of errors per hundred words. A detailed subclassification of the errors will be performed so that error frequency may be established and the most characteristic types of errors found. Correlations will be calculated between the most salient error types and the error rates to establish whether particular error types have a predictive power for proficiency. The following hypotheses and research questions will be considered.

Research questions — accuracy

1. Can the most problematic areas (e.g. persistent errors) be identified?

2. Are problematic areas linked to particular tasks?

3. Is the group of students examined in our research homogenous? How much variance in proficiency is there? Does the same dispersion apply to all of the tasks?

4. Are lower-level students characterized by errors which higher-level students do not make?

5. Can a higher error rate predict particular types of errors?

6. Can the sources of most frequently occurring errors be identified?

Hypotheses — accuracy

1. Grammatical errors are the most frequent error type.

2. Lexical errors are the second most frequent group error type.

3. Errors in the use of articles are the most frequent type of grammatical errors.

4. Errors in the use of tenses are the second most frequent type of grammatical errors.

5. Errors in the use of lexical phrases are the most frequent type of lexical errors.

6. Errors in the use of prepositions are the second most frequent type of lexical errors.

7. There is a correlation between errors in the use of articles and error rate.

8. There is a correlation between lexical errors and error rate.

9. There is a correlation between errors in the use of lexical phrases and error rate.

10. There is a correlation between errors in the use of tenses and error rate.

11. Task one has the strongest effect on accuracy.

12. Task three has the weakest effect on accuracy.

3.1.3 Pilot study — fluency

For the purposes of the pilot study fluency was operationalized as speed fluency and breakdown fluency (Skehan 2003; Tavakoli and Skehan 2005). To this end, speech rates were measured and pause phenomena evaluated (frequency of unfilled (UP) and filled (FP) pauses, mean length of runs (MLR)). As previous research has shown (e.g. Robinson 2001; Ellis and Yuan 2004; Levkina and Gilabert 2012) task variability plays a considerable role in fluency measurements. I have therefore calculated the selected variables separately for all of the three taks so that comparisons can be made.

3.1.3.1 Pilot study — speech rate

Speech rate was measured as words per minute. I considered using the syllables per minute rate as this is more accurate. I experimented with online syllable counters but these proved highly inaccurate and unreliable. Counting syllables manually, however, proved rather time consuming. To check how much precision I would lose in selecting words per minute rather than syllables per minute I carried out a simple experiment. I chose random passages of text of comparable length produced by the six speakers in the pilot study and calculated the ratio of syllables per words. Table 3–7 shows the comparisons.

Table 3–7 Pilot study — fluency — syllable-word ratios in a selection of random passages by the pilot speakers

	PS1	PS2	PS3	PS4	PS5	PS6	Mean	SD
Word count	287	320	315	305	332	296	**309**	15.3
Syllable count	373	400	419	372	422	400	**398**	19.69
Syllable/word ratio	1.30	1.25	1.33	1.22	1.27	1.35	**1.29**	.05

The table shows that the speakers produced between 1.22 and 1.35 syllables per word. The average length of a word is 1.29 syllables. The very low standard deviation of .05 shows that the

differences in word length produced by the individual speakers is negligible. The syllable count is simply a multiple of the word count. It is therefore not necessary for the purposes of our study to express speech rate in syllables per minute. Instead, words per minute, which is much less time-consuming to calculate, can be used.

Measuring speech rate proved to be rather laborious. Even in tasks one and three, which are designed as primarily monological, there appear to be interruptions by the interviewer both in the forms of simple backchannelling and as verbalized reactions or questions. This precluded the simple option of working out the speech rate as a ratio of words uttered and time taken up by the whole production. Speech rate studies rarely mention the exact measuring procedure. Lennon (1990a) mentions timing the recording with a stopwatch, which is however unreliable at the fast pace of largely dialogic interviews with overlaps, and frequent interruptions. Consequently, I decided to use an audio-editing package (Wavepad) and manually cut out all utterances (except overlaps) that were made by the interviewer. This is a time-consuming technique which however yields reliable results as it does not happen in real time (the recording can be stopped, the researcher can reassess passages etc.) The resulting audio track thus contained only utterances made by the students. I used the Wavepad's bookmarking facility to mark boundaries between the tasks. The length of the tasks was then entered in a spreadsheet along with the respective word counts (I used unpruned words only). Fig. 3–9 shows the resulting speech rates.

Table 3–8 Pilot study — speech rates in words per minute (non-native speakers)

Speaker	PS1	PS2	PS3	PS4	PS5	PS6	Mean	SD
Task 1	206	137	199	135	167	172	**169.33**	29.87
Task 2	198	139	204	147	173	189	**175.00**	27.02
Task 3	182	118	190	121	146	142	**149.83**	30.23

As the table reveals, speakers PS1 (selected as the most fluent speaker) and PS3 (selected as perceptively the fastest) have the highest speech rates in all of the tasks. Speakers PS3 (selected as the weakest) and PS4 (perceptively evaluated as the slowest) have the lowest speech rates. Perceptive evaluation of the speakers matches the results obtained by the measurements.

A one-way repeated measures ANOVA was carried out to compare speech rates in the three tasks. The means and standard deviations are presented in Table 3–8 above. There was a significant effect for task variability, Wilks' Lambda = .105, F (2, 4) = 17.06, $p < .05$. The multivariate partial eta squared of 0.895 suggests a very large effect size. To determine which of the differences is statistically significant all of the tasks were compared in a paired-samples t-test. A statistically significant difference was found between Tasks 1 and 3, t (5) = 6.46, $p < .05$ (2-tailed). The partial eta squared of .81 indicated a very large effect size. A statistically significant difference was also found between Tasks 2 and 3, t (5) = 5.19, $p < .05$ (2-tailed) with a very large effect size (partial eta squared = .73). Comparisons between Tasks 1 and 2 did not yield statistically significant results ($p > .05$).These results suggest that the three tasks in our dataset ought to be treated separately as task variability has a large effect on the speakers' performance.

3.1.3.2 Pilot study — unfilled pauses (UPs)

The identification of UPs was facilitated by the transcriptions, where pauses are marked by full stops, distinguishing between short (<1s), medium (1–3s) and long pauses (>3s). Following Götz's (2013) approach I made no distinction between them as pause length has not been found to have a significant effect on learners' fluency (Götz 2013, Cucchiarini et al. 2002). I counted pauses and calculated a pause rate (pauses per hundred words) in the three tasks separately using a computer script I wrote for this purpose. Table 3–9 shows the results.

Table 3–9 Pilot study — fluency — frequency of unfilled pauses per hundred words (non-native speakers)

	PS1	**PS2**	**PS3**	**PS4**	**PS5**	**PS6**	**Mean**	**SD**
Task 1	8.9	15.2	12.0	6.8	13.8	8.7	**10.9**	3.3
Task 2	7.14	12.11	11.20	7.28	10.40	6.42	**9.1**	2.4
Task 3	13.95	15.69	11.70	10.85	18.06	11.24	**13.6**	2.9
Tasks 1—3	8.87	13.52	11.57	7.93	12.31	8.77	**10.41**	2.3

The table shows that for all of the speakers the frequency of UPs is largest in task 3. No previously made conclusions regarding the perceived fluency or speed seem to apply to pause rate. The slowest speaker PS4 has the lowest UP frequency, whilst the fastest speaker the third highest.

3.1.3.3 Pilot study — filled pauses (FPs)

The filled pauses in the transcriptions were counted in two steps. First of all, pauses which indicate backchannelling were deleted. This had to be evaluated manually along with listening to the transcriptions. Subsequently, the remaining FPs were counted automatically using my own computer script and a frequency per hundred words was calculated as shown in Table 3–10.

Table 3–10 Pilot study — fluency — frequency of filled pauses per hundred words (non-native speakers)

	PS1	PS2	PS3	PS4	PS5	PS6	Mean	SD
Task 1	4.86	8.61	8.77	9.40	3.13	2.98	**6.29**	2.97
Task 2	4.29	9.78	9.21	6.65	3.08	2.03	**5.84**	3.23
Task 3	5.12	11.38	11.92	8.79	3.70	2.07	**7.16**	4.13
Tasks 1-3	4.76	9.70	9.64	7.82	3.17	2.71	**6.30**	3.17

The table shows that Task 3 has the highest frequency of FPs for all of the speakers. The dispersion of the values is considerable (e.g. the mean for Tasks 1—3 is 6.3, SD = 3.17), as is the range (from 2.07 to 11.92). The highest frequencies of FPs were found for speakers PS2 (weakest), PS3 (fastest) and PS4 (slowest).

3.1.3.4 Pilot study — mean length of runs

A length of run is the number of words uttered between two pauses (Lennon 1990). To identify runs automatically I wrote a computer script which placed each run on a separate line and provided the word count (of unpruned words) in a separate column. Subsequently, I rejected runs which were short (e.g. 2- or 3-word) answers to the interviewer's questions, and runs which were cut-off by the interviewer. The resulting text file was then imported into a spreadsheet and the mean length of runs was calculated for each individual task. The results are shown in Table 3–11. The procedure of calculating the length of runs is a markedly laborious task as all of the interviewer's utterances have to be evaluated independently as to whether they present an interruption or not. Sometimes it is even necessary to listen to the recording to establish whether an utterance was terminated by the learner or by the interviewer's interruption.

Table 3–11 Pilot study — fluency — mean length of runs in words (non-native speakers)

	PS1	PS2	PS3	PS4	PS5	PS6	Mean	SD
Task 1	10.63	6.5	7.32	11.1	6.83	8.58	**8.49**	1.97
Task 2	10.76	6.08	6.76	7.82	6.46	8.44	**7.72**	1.73
Task 3	7.17	5.87	7.53	5.62	5.38	6.34	**6.32**	0.87
Tasks 1—3	9.52	6.15	7.20	8.18	6.22	7.79	**7.51**	1.28

The speaker perceived as most fluent (PS1) produced the longest runs, unlike the weakest speaker whose runs were the shortest. All speakers produced their shortest runs in Task 3.

3.1.3.5 Pilot study — native speakers' speech rate, frequency of UPs and FPs, and mean length of runs

Six native speakers were selected for the pilot study. The only criteria which were identical to those used for the selection of the learners were perceptive speech rate. By listening to brief extract I tried to choose the fastest and the slowest speaker. I also chose a speaker who appeared to use a large number of filled pauses. As I could not apply accuracy in the NS speech as one of the defining criteria as I did with the learners, the remaining three speakers were selected randomly.

Speaker NS1 — perceptively fastest;
Speaker NS2 — perceptively slowest;
Speaker NS3 — perceptively frequent use of filled pauses;
Speakers NS4–NS6 — randomly selected.

Using the techniques described above, for each task I calculated their speech rates (Table 3–12), frequencies of UPs and FPs (Tables 3–13 and 3–14), and the mean lengths of runs (Table 3–15).

Table 3–12 Pilot study — speech rates in words per minute (native speakers)

	NS1	NS2	NS3	NS4	NS5	NS6	Mean	SD
Task 1	243	155	195	221	202	212	**204.86**	29.63
Task 2	242	183	190	227	215	221	**213.12**	22.52
Task 3	193	128	142	158	183	182	**164.35**	25.85

Table 3–13 Pilot study — fluency — frequency of UPs per hundred words (native speakers)

	NS1	NS2	NS3	NS4	NS5	NS6	Mean	SD
Task 1	3.87	7.49	5.65	4.11	2.60	6.33	**5.01**	1.80
Task 2	3.70	6.74	6.48	2.65	1.58	5.83	**4.50**	2.16
Task 3	10.34	14.71	13.40	8.06	1.49	8.37	**9.40**	4.70
Tasks 1-3	4.13	7.39	6.19	3.72	1.89	6.20	**4.92**	2.03

Table 3–14 Pilot study — fluency — frequency of FPs per hundred words (native speakers)

	NS1	NS2	NS3	NS4	NS5	NS6	Mean	SD
Task 1	1.17	4.56	2.88	.85	2.38	1.79	**2.27**	1.35
Task 2	1.14	3.45	2.02	1.02	1.90	.70	**1.70**	1.00
Task 3	2.59	6.86	3.09	1.61	2.69	3.94	**3.46**	1.83
Tasks 1-3	1.23	3.88	2.60	.96	2.14	1.29	**2.02**	1.11

Table 3–15 Pilot study — fluency — mean length of runs in words (native speakers)

	NS1	NS2	NS3	NS4	NS5	NS6	Mean	SD
Task 1	8.5	6.84	8.82	7.62	14.38	9.44	**9.27**	2.67
Task 2	8.47	7.01	7.43	7.9	15.2	8.45	**9.08**	3.05
Task 3	8.83	5.1	5.9	8.71	23.64	8.2	**10.06**	6.83
Tasks 1-3	8.60	6.32	7.38	8.08	17.74	8.70	**9.47**	4.15

The native-speaker speech variables appear to be affected by task variability in the same way as in the case of the learners. Task 3 on the whole proves to be most complex, yielding the slowest speech rate, the highest number of UPs and FPs, and the shortest runs. In comparing native and non-native performance I will therefore consider each task separately.

As regards speech rate, the native speakers were faster in all three tasks. The boxplots in Fig. 3–2 show a separate boxplot for each task. The first three boxplots illustrate the performance in the three tasks by the 6 learners, the last three boxplots those by the 6 native speakers. T-tests comparing the three pairs of tasks showed significant statistical differences for Task 1 ($t(5) = -2.9$, $p < .05$) and Task 2 ($t(5) = -3.07$, $p < 0.05$). The difference for Task 3 ($t(5) = -1.06$, $p > .05$) is not statistically significant.

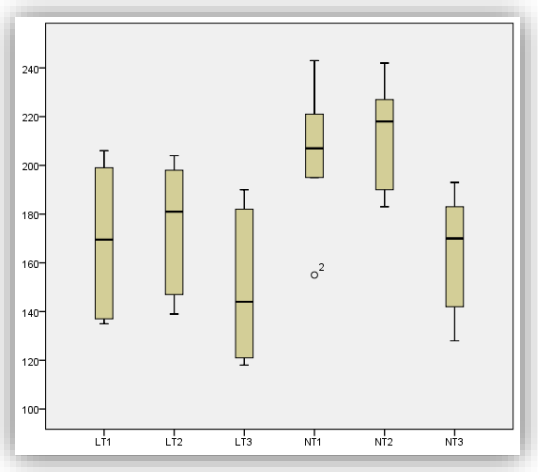

Fig. 3–2 Boxplots comparing speech rates in the three tasks (T1, T2, T3) for learners (L) and native speakers (N). The y axis represents speech rate in words per minute.

As regards the frequency of UPs (see Fig. 3–3), the log-likelihood values ($G2 > 15.13$, $p < .0001$) show that in all three tasks the learners overuse UPs.

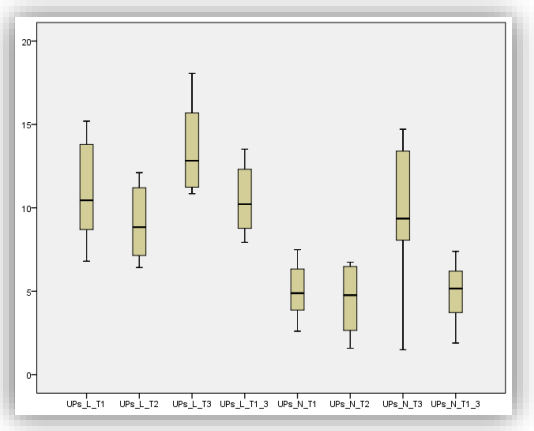

Fig. 3–3 Boxplots comparing frequency of UPs in the three tasks (T1, T2, T3) for learners (L) and native speakers (N). The y axis represents unfilled-pause rate as number of pauses per hundred words.

As regards the frequency of FPs (see Fig. 3–4), the log-likelihood values ($G2 > 15.13$, $p < .0001$) show that in all three tasks the learners overuse FPs.

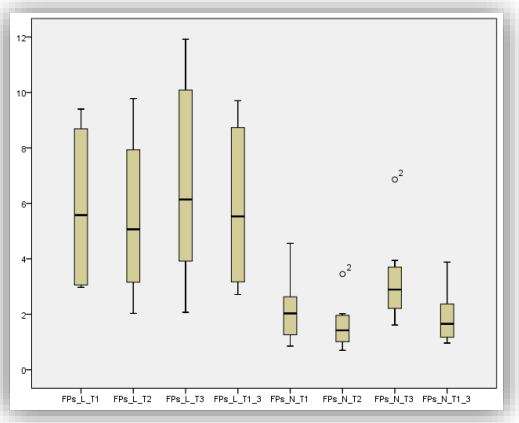

Fig. 3–4 Boxplots comparing frequency of FPs in the three tasks (T1, T2, T3) for learners (L) and native speakers (N). The y axis represents filled-pause rate as number of filled pauses per hundred words.

As far as the length of runs is concerned (see Fig. 3–5), native speakers in the pilot study produce longer runs but within the small sample there do not appear to be significant differences between any of the tasks ($p > .05$).

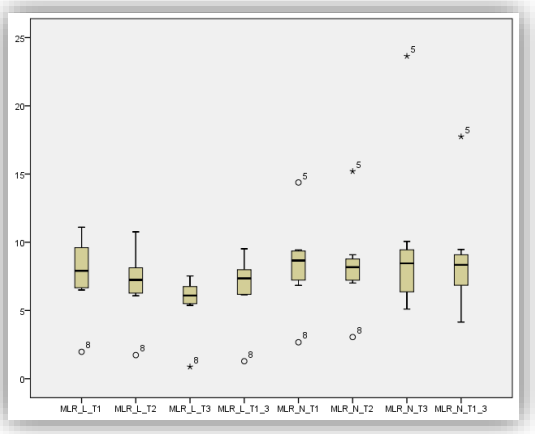

Fig. 3–5 — Boxplots comparing MLRs in the three tasks (T1, T2, T3) for learners (L) and native speakers (N). The y axis represents mean length of runs in words.

3.1.4 Pilot study — fluency — conclusion

The pilot study for fluency has proved the feasibility of the selected techniques provided these can be partly automated using computer scripts. Whilst these somewhat reduce the laboriousness of the tasks, interventions by the researcher are necessary as many of the features (e.g. the function of the FPs, or the termination of a run) can only be accurately evaluated along with listening to the recordings. Measuring the features of fluency in a large corpus is a markedly time-consuming task.

The results show that as with accuracy there is a high dispersion in the data. The comparison between learners and native speakers shows lower performance in all aspects on the part of the learners. Task variability proves to have a significant effect on the performance and it is thus essential that all measurements be carried out separately in the individual tasks and not across the three tasks considered together as one text.

Owing to the small scale of the pilot study it is not possible to reach significant results of correlations. In the final study, however, these will be evaluated using Pearson's r, trying to establish whether correlations exist between fluency and accuracy in the individual tasks.

The following hypotheses will be tested in the final study:

1. Task design affects both native- and non-native-speaker fluency.

2. In all three tasks the learners produce fewer words per minute than the native speakers.

3. In all three tasks the learners overuse UPs.

4. In all three tasks the learners overuse FPs.

5. In all three tasks the learners produce shorter speech runs than the native speakers.

6. There is a correlation between accuracy and speed fluency.

7. There is a correlation between accuracy and breakdown fluency.

4. Accuracy — results and analyses

The following section presents the results of accuracy measurements in LINDSEI_CZ. As shown in the pilot study, accuracy is operationalized as error rate, expressed as the normalized frequency of errors per hundred words. The chapter deals, first of all, with a quantitative error analysis of the individual error types, and then it provides a presentation of error rates and identification of persistent errors.

4.1 Error analysis

The results presented in this chapter follow a process of detailed error analysis and tagging whose principles were described in section 2.4.1 of this thesis. A total of 1,299 errors were identified and divided into the six subgroups as described in the Louvain error-tagging manual. Table 4–1 shows a summary of error frequencies in the subgroups in the individual tasks, and in the last two columns the totals for all tasks.

Table 4–1 Frequencies of errors in Tasks 1, 2 and 3, and in all tasks together

	Task 1		Task 2		Task 3		All tasks	
	Count	%	Count	%	Count	%	Count	%
Morphological errors (inflectional, derivational)	0	**.0%**	3	**.5%**	0	**.0%**	3	**.23%**
Grammatical errors	328	**61.1%**	327	**55.6%**	55	**31.6%**	710	**54.66%**
Lexico-grammatical errors	24	**4.5%**	34	**5.8%**	13	**7.5%**	71	**5.47%**
Lexical errors	146	**27.2%**	195	**33.2%**	93	**53.4%**	434	**33.41%**
Word redundant/missing, word order errors	26	**4.8%**	18	**3.1%**	7	**4.0%**	51	**3.93%**
Infelicities	13	**2.4%**	11	**1.9%**	6	**3.4%**	30	**2.31%**
Total	**537**	100.0%	**588**	100.0%	**174**	100.0%	**1,299**	100.00%

These results confirm the initial findings of the pilot study which posited that grammatical errors are more frequent than lexical errors but that the two groups form the largest proportion of errors in the whole dataset. Within tasks, the results are similar, except the reversed frequencies for lexical and grammatical errors in Task 3. These findings are contrary to claims (e.g. Agustín Llach 2011) that lexical errors are the most frequent ones. The two groups form 88% of all of the errors in LINDSEI_CZ and are committed by practically every speaker. The only other group which shows

a frequency of errors of more than 5% is the lexico-grammatical group (5.47%) but its low count along with the low frequency in the other groups seem to suggest that the errors in these groups are rather non-systematic.

4.1.1 Form errors (the F* group)

In the Louvain error-tagging system the F* tag is used to label all non-existent words. At the same time a distinction is made between errors in the spelling (tagged FS) and morphological errors (FM), which can be of derivational nature (use of inappropriate affix) or inflectional (misuse of grammatical morphemes). Although the students who participated in LINDSEI_CZ transcribed their own recordings, spelling was not evaluated as LINDSEI is designed as a spoken corpus. Also many of the possible instances of the morphological errors described in the manual would only apply to writing (e.g. the use of apostrophe). As a result, only three FM errors were identified in the whole dataset (**pedagogian, *psychologian, *realias*).

4.1.2 Grammatical errors (the G* group)

A grammatical error is the result of a breach of the rules of English grammar. As described in the pilot study, all errors were tagged unless they were immediately followed by a self-correction, or unless they were immediately repeated. A total of 710 grammatical errors (54.66%) were identified, which shows that grammar is the most error-prone area of our learners' use of English. With the exception of speaker CZ023 every learner committed at least one grammar error. Fig. 4–1 shows the grammar error counts for individual speakers.

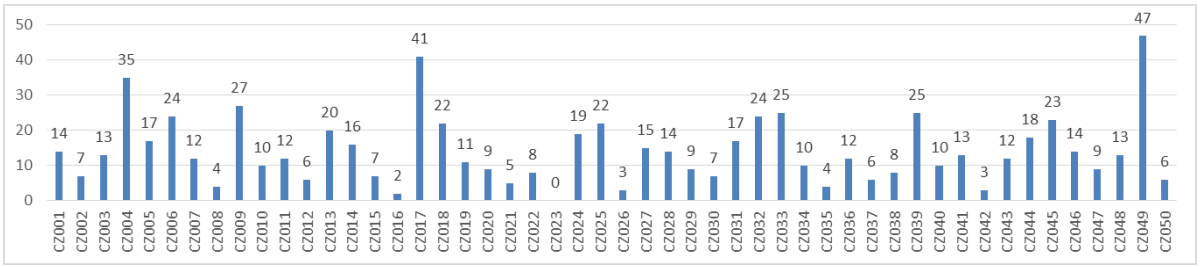

Fig. 4–1 Grammar error (G) counts per speaker. The numbers above the bars denote the number of grammatical errors.*

The Louvain error-tagging manual recognizes 8 groups of grammatical errors most of which are subdivided into subgroups. Altogether it offers 25 different tags for grammar errors. The 8 groups include determiners (tagged GD), articles (GA), nouns (GN), pronouns (GP), adjectives (GADJ), adverbs (GADV), verbs (GV), and word class (GWC). As a result of my own error analysis of LINDSEI_CZ I added 3 groups: numerals (GNUM), existential construction (GEX) and proforms (GPRO). I also made some additions to the subgroups, which are discussed in the relevant sections below. Table 4–2 shows the basic descriptive statistics.

Table 4–2 Grammatical errors — absolute and relative frequencies (n = number of erring speakers)

Type	**Count**	**%**	**n**
Determiners (GD)	44	6.20%	22
Articles (GA)	265	37.32%	44
Nouns (GN)	26	3.66%	19
Pronouns (GP)	44	6.20%	25
Adjectives (GADJ)	4	.56%	4
Adverbs (GADV)	52	7.32%	28
Numerals (GNUM)	5	.70%	5
Verbs (GV)	249	35.07%	46
Word class (GWC)	15	2.11%	11
Existential construction (GEX)	2	.28%	2
Proform (GPRO)	4	.56%	4
Total	710	100.00%	

As the table shows, the use articles (37.32%) and verbs (35.07%) present the biggest problems. Besides that it is errors in the use of adverbs (7.32%), determiners (6.2%), pronouns (6.2%), and nouns (3.66%). The remaining groups are sparsely populated. The following section provides a more detailed analysis of the individual groups.

4.1.2.1 Errors in the use of determiners (the GD group)

This category subsumes all determiner errors except those in the use of articles. Four subcategories are distinguished:

– demonstrative determiners (GDD) — *with (GDD) this $these$ studies;*

– possessive determiners (GDO) — *at (GDO) hers her place;*

– indefinite determiners (GDI) — *it started (GDI) few $a few$ years ago;*

– other types of determiner (GDT), including predeterminers and postdeterminers — *they are (GDT) so $such$ great books*.

Table 4–3 Determiner errors — absolute and relative frequencies (n = number of erring speakers)

Type	Count	%	n
Demonstrative	1	2.27%	1
Possessive	4	9.09%	4
Indefinite	31	70.45%	18
Other	8	18.18%	7
Total	**44**	**100.00%**	

Table 4–3 shows the breakdown of results in the individual groups of determiner errors. This reveals that the most problematic group in this part is the use of indefinite determiners. A close analysis of these errors revealed a high number of instances of wrongly used quantifiers, such as the use of *lot of* instead of *a lot of/lots of*, *few* instead of *a few*, and *much* instead of *many*. However, with regard to the low relative frequency of these errors and the low number of students who committed them, these errors can be considered marginal.

4.1.2.2 Errors in the use of articles

As was shown above in Table 4–2 article errors are among the most frequent, forming 37.32% of all grammar errors, and 20.4% of all errors in LINDSEI_CZ. At least one article error was committed by 44 (88%) speakers. Article errors are tagged (GA) and are not divided into subgroups. Thus it is rather laborious to determine which aspects of their use are most problematic. A more detailed analysis (see Table 4–4) revealed that 56.6% of all of the errors involved an erroneous omission of either the indefinite (34.34%) or the definite (22.26%) article. Such a result would seem to suggest that as Czech does not have articles, article omission in English might be a result of L1 transfer. The third largest group (27.17%) was made up of instances in which a definite article was used instead of a zero article. The results show that the use of articles is highly problematic even for advanced learners.

Table 4–4 Types and frequencies of article errors (the sign 0 stands for article omission)

Type	Count	%
(GA) 0 a/an	91	34.34%
(GA) 0 the	59	22.26%
(GA) the 0	72	27.17%
(GA) the a/an	13	4.91%
(GA) a/an 0	16	6.04%
(GA) a/an the	4	1.51%
(GA) the $possessive pronoun$	3	1.13%
(GA) a/an $possessive pronoun$	3	1.13%
(GA) a an	2	.75%
(GA) the $some$	1	.38%
(GA) one a	1	.38%
Total	**265**	**100%**

The bar chart in Fig. 4–2 displays article-error counts for the individual speakers, facilitating their comparison: 6 speakers did not commit any article errors, 23 speakers commited 1–5 article errors, 13 speakers committed 6–10 errors, and 8 speakers commited more than 10 article errors.

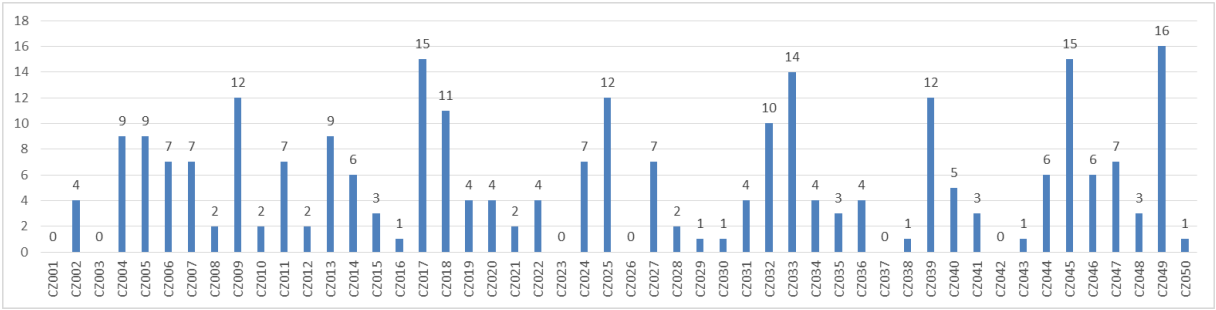

Fig. 4–2 Article-error counts for all LINDSEI_CZ speakers

4.1.2.3 Errors in the use of nouns

Noun errors are tagged as (GN) and two subgroups are distinguished:
– noun number (GNN) — *one of my (GNN) visit $visits$ was (er) quite a long one;*
– noun case (GNC) — *to study for a master $master's$ degree.*

Table 4–5 Noun errors — absolute and relative frequencies (n = number of erring speakers)

Type	Count	%	n
Noun number	25	96.15%	18
Noun case	1	3.85%	1
Total	**26**	**100.00%**	

Noun-number errors are frequently errors in concord between subject and object (*they have to spend their whole (GNN) life $lives$ there*), which is a typical problem for Czech learners as in Czech

the object in such examples is not affected by concord rules. Noun case errors are infrequent, and the one example listed above could also be classified as a lexical single collocation error (LS). More genitive errors in the use of the apostrophe could be expected in writing, but such errors are obviously undetectable in speech.

4.1.2.4 Errors in the use of pronouns

Pronominal errors are tagged (GP) and are divided into the following subgroups:

– demonstrative pronouns (GPD) — *I like translating too much to not to do (GPD) that it;*

– personal pronouns (GPP) — *if (GPP) it he was an artists;*

– possessive pronouns (GPO) — *friend of (GPO) her $hers$;*

– indefinite pronouns (GPI) — *didn't have any mobile phones or (GPI) something $anything$ like that;*

– reflexive and reciprocal pronouns (GPF) — (no example in LINDSEI_CZ);

– relative and interrogative pronouns (GPR) — *the actors (GPR) which who are really good;*

– unclear pronominal reference (GPU) — (no example in LINDSEI_CZ).

Table 4–6 Pronoun errors – absolute and relative frequencies (n = number of erring speakers)

Type	Count	%	n
Demonstrative pronouns (GPD)	4	9.09%	4
Personal pronouns (GPP)	21	47.73%	18
Possessive pronouns (GPO)	2	4.55%	2
Indefinite pronouns (GPI)	2	4.55%	1
Relative and interrogative pronouns (GPR)	15	34.09%	10
Total	**44**	100.00%	

The two largest groups are those involving personal and relative/interrogative pronouns. Errors in the use of personal pronouns include especially examples of pronoun omission (*the weather got quite terrible and (GPP) 0 it started raining*), and those in which the pronoun *we* is used instead of the singular *I ((GPP) we I went there with my friends*), which can be considered as an example of L1 transfer. As for relative pronouns, there are a number of examples where *which* is used instead of *who* or vice versa.

4.1.2.5 Errors in the use of adjectives

Only four adjectival errors were identified in LINDSEI_CZ. They are tagged (GADJ), and fall into three subgroups:

– adjective order (GADJO) — (no example in LINDSEI_CZ);

– adjective number (GADJN) — *the (GADJN) olders $older ones$ are eleven;*

– comparative/superlative (GADJCS) — *speak about my (GADJCS) most favourite $favourite$ movie.*

4.1.2.6 Errors in the use of adverbs

Adverbial errors are tagged (GADV). The Louvain manual distinguishes only one type, namely the adverbial word order error (GADVO). I have also made use of the tag (GADV) to label general adverbial errors, and introduced a new tag (GADVI) to label errors in the use of intensifiers.

– general adverbs error (GADV) — *I know (GADV) lot $a lot$ about (eh) the real;*

– adverbial word order (GADVO) — *there was (GADVO) a band playing also $there was also a band playing$;*

– intensifiers (GADVI) — *it's also (GADVI) little bit $a little bit$ weird.*

Table 4–7 Adverb errors — absolute and relative frequencies (n = number of erring speakers)

Type	**Count**	**%**	**n**
General	7	13.46%	6
Adverbial word order	27	51.92%	18
Intensifiers	18	34.62%	13
Total	**52**	**100.00%**	

4.1.2.7 Errors in the use of verbs

Along with articles, errors in the category of verbs are the most frequent, forming 35.07% of all grammar errors and 19.17% of all errors in LINDSEI_CZ. At least one grammatical error was committed by 46 (92%) speakers. Verbal errors are tagged (GV) and in the Louvain manual they are split into 6 subgroups to which I added the tags (GVMC), (GVTA), (GVTC) and (GVAUXC):

– verb number (GVN) — *her hair (GVN) are is different;*

– verb morphology (GVM) — *it was very interesting for me to (GVM) found $find$ out;*

– verb morphology — conditional (GVMC) — *I think that they would (GVMC) kill $have killed$ us;*

– non-finite/finite verb forms (GVNF) — (no example in LINDSEI_CZ);

– verb voice (GVV) — *the children are running to (eh) (GVV) get hidden $hide$ somewhere in the garden;*

– verb tense and aspect (GVT) — *and the whole time (GVT) it hasn't rained $it didn't rain$;*

– verb tense agreement (GVTA) — *he told us that (GVTA) it's $it was$ raining everywhere in Wales as well;*

– verb tense conditional (GVTC) — *if they (GVTC) stayed $had stayed$ over there.*

– auxiliaries (GVAUX) — *it was near the ferry so we (GVAUX) can $could$ take an early ferry*;

– auxiliaries — conditional (GVAUXC) — *if it (GVAUXC) would was a close friend I would say something.*

Table 4–8 Verb errors — absolute and relative frequencies (n = number of erring speakers)

Type	Count	%	n
Verb number (GVN)	14	5.62%	10
Verb morphology (GVM)	20	8.03%	11
Verb morphology — conditional (GVMC)	7	2.81%	6
Verb voice (GVV)	2	0.80%	2
Verb tense (GVT)	128	51.41%	39
Verb tense agreement (GVTA)	44	17.67%	22
Verb tense conditional (GVTC)	5	2.01%	4
Auxiliaries (GVAUX)	18	7.23%	15
Auxiliaries — conditional (GVAUXC)	11	4.42%	8
Total	**249**	**100.00%**	

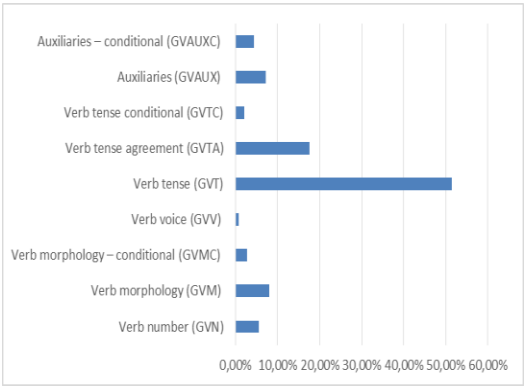

Fig. 4–3 Frequency of verb errors

As shown in Table 4–8, errors in the use of tense and aspect are the most frequent ones. They have been split here into three subgroups to distinguish errors in tense agreements and use of tenses in conditional clauses. The three groups together present 71.1% of all verb errors, 24.9% of all grammar errors, and 13.6% of all errors in LINDSEI_CZ. The bar chart in Fig. 4–4 displays verb-error counts for the individual speakers, facilitating their comparison: 8 speakers did not commit any tense errors, 30 speakers committed 1–5 tense errors, 10 speakers committed 6–10 errors, and 2 speakers committed more than 10 tense errors. These figures show that tenses are clearly highly problematic even for advanced learners.

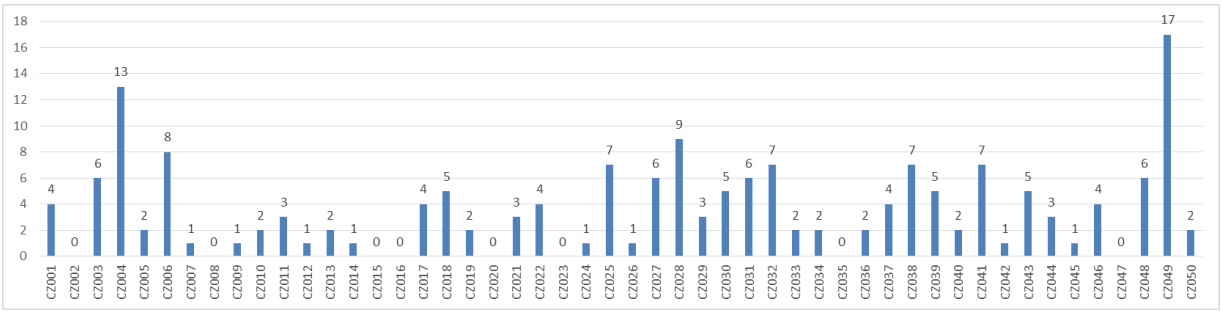

Fig. 4–4 Verb-error counts for all LINDSEI_CZ speakers

To obtain a deeper understanding of the (GVT) group of errors (not including agreement and conditional errors) I analyzed the 128 tense errors in detail, trying to identify patterns of misuse. The results are shown in Table 4–9.

Table 4–9 Types of verb tense and aspect errors in LINDSEI_CZ

Type	**Count**	**%**
Past simple for present perfect	50	39.10%
Present perfect for past	24	18.80%
Past simple for past perfect	15	12.03%
Present for past	12	9.02%
Present perfect for past perfect	10	7.52%
Present simple for present perfect	7	5.26%
Past for present	3	2.26%
Expressing the future	2	1.50%
Past perfect for present perfect	1	.75%
Present perfect for present simple	1	.75%
Present continuous for present simple	1	.75%
Past continuous for past simple	1	.75%
Present simple for past perfect	1	.75%
Present simple for present continuous	1	.75%
Total	**128**	**100%**

The table shows that the most frequent type of tense error (39.1%) involves the use of the past simple instead of the present perfect. All in all, errors involving the present perfect form 71.9% of all tense errors. A closer look at the errors reveals that of all present perfect errors 38.8% are formed by the erroneous use of the past simple instead of the present perfect, and 18.7% by the erroneous use of the present perfect instead of the past simple. Whilst we could conclude that the present perfect is the hardest tense for Czech advanced learners we must also point out that the high frequency of these errors might be affected in LINDSEI_CZ by the design of the tasks. In Task 1 the learners are expected to talk about a past experience which has had a lasting effect on them. Such topics invite the use of the present perfect. Similarly, in the interview in Task 2 many of the

interactions revolve around the students' experience, the questions are often about the way they started studying English, why they chose to study linguistics etc. As in Task 1, these situations require a frequent use of the past tense but often also the present perfect. The constant tension between the two tenses results in tense misuse. The design of these tasks thus very successfully manages to identify one of the greatest weaknesses in our learners' command of English.

4.1.2.8 Errors in the use of word class (the GWC group)

This group includes all examples in which an inappropriate word class is used. In LINDSEI_CZ there were only 15 such instances, amongst which it possible to find only one type of repeated error, which is the use of an adverb instead of an adjective following a copular verb. This is undoubtedly due to L1 transfer.

that don't really sound all that (GWC) well $good$ in Czech (cf. the Czech translation: *které v češtině nezní moc dobře*)

4.1.2.9 Errors in the use of the existential construction (the GEX group)

This is a small group of errors which could not be classified using the existing Louvain manual which does not have a separate tag for these errors. Consequently, I introduced a new tag (GEX). However, only two instances of the error were identified in LINDSEI_CZ.

4.1.2.10 Errors in the use of proform (the GPRO group)

The Louvain manual makes no allowance for tagging errors in the use of proforms. Consequently I introduced the (GPRO) tag. Only four of these errors were identified in LINDSEI_CZ.

4.1.3 Lexico-grammatical errors (the X* group)

As its name suggests, this group is on the boundary between grammatical and lexical errors. In order to avoid committing such an error, the speaker not only has to make the right lexical choice but he also has to pay attention to its morpho-syntactic properties. Typically, these errors involve

complementation, dependent prepositions, and countability. As regards complementation errors, three types were identified in our data:

– noun complementation (XNCO) — *"it was my own (XNCO) fault to go $fault that I went$ there";*

– adjectival complementation (XADJCO) — *"here I am (XADJCO) used to work $used to working$ with English";*

– verbal complementation (XVCO) — *"firstly I was (XVCO) thinking to go $thinking of going$ on Erasmus".*

There also appeared to be three types of errors including dependent prepositions:

– nouns followed by a wrong preposition (XNPR) — *"we had a gorgeous (XNPR) view on $view of$ the Himalayas;*

– adjectives followed by a wrong preposition (XADJPR) — *"I've never been (XADJPR) interested into $interested in$ food";*

– verbs followed by a wrong preposition (XVPR) — *"it (XVPR) reminds me $reminds me of$ Oscar Wilde"*

The X* group also includes errors in countability (XNUC). These were very infrequent in our data:

– noun countability (XNUC) — *"because she has straight (XNUC) hairs $hair$ over there"*

Lexico-grammatical errors form 5.41% of all errors identified in LINDSEI_CZ and thus do not appear to be a particularly problematic area for our speakers. These errors were not committed by all of the speakers (see Fig. 4–5).

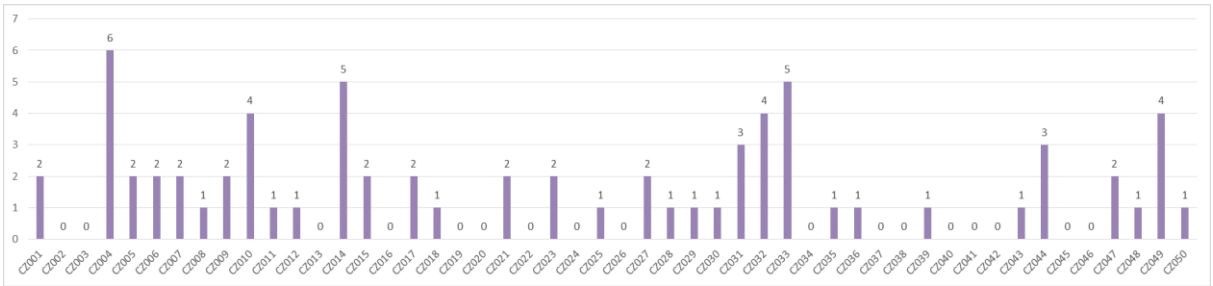

Fig. 4–5 Lexico-grammatical error (X) count per speaker*

Table 4–10 Lexico-grammatical errors — absolute and relative frequencies (n = number of erring speakers)

Type	Count	%	n
Complementation errors	25	35%	16
Dependent prepositional errors	44	62%	28
Countability errors	2	3%	2
Total	**71**	**100.00%**	

Table 4–11 Complementation errors — absolute and relative frequencies (n = number of erring speakers)

Type	Count	%	n
XADJCO	5	20.0%	3
XNCO	3	12.0%	3
XVCO	17	68.0%	14
Total	**25**	**100.0%**	

Table 4–12 Dependent-preposition errors — absolute and relative frequencies (n = number of erring speakers)

Type	Count	%	n
XADJPR	9	20.5%	8
XNPR	11	25.0%	8
XVPR	24	54.5%	19
Total	**44**	**100.0%**	

As shown in Tables 4–10—4–12 the group of lexico-grammatical errors contains mostly errors in the use of dependent prepositions (28 speakers committed at least one error) and especially prepositional verbs (19 speakers committed at least one error). Complementation errors were less frequent (affecting 16 speakers) and especially in the use of verbal complementation (committed by 14 speakers). The low frequency of all of the types within the group suggests that these errors are of random nature and general conclusions are difficult to make. A close look at the errors reveals the presence of some stereotypical problems regularly described in literature on errors (e.g. **arrive to, *explain sb, *listen sth, *be used to + infinitive, *make sb to do sth*). L1 transfer is sometimes detectable (e.g. **blind on one eye, *view on the Himalayas, *books from an author, *exam from linguistics, *see into so's face, *write with each other, *graduate from maths*).

4.1.4 Lexical errors (the L* group)

The Louvain tagging system[30] distinguishes between lexical single errors (affecting single words), lexical phrases (affecting multiple words), and lexical connectors. It is based on a simple, product-oriented descriptive classification which includes eight groups. Errors "[involve] the semantic (conceptual, collocational, or connotative) properties of words or phrases" (Dagneaux et al.,

[30]The weakness of the Louvain error classification is in the very narrow range of lexical errors.

2008, 30) but these concepts are not explicitly defined, only exemplified. There is no explicit definition of an error or a word. Lexical errors are simply set against any of the other categories, especially that of grammatical errors. A lexical error is a result of a wrong selection of a lexical item (cf. Lennon, 1991) but it does not stem from ignorance of morpho-syntactic properties of any of the items involved. In this respect it is different from the lexico-grammatical group of errors. Form-oriented errors are not included as they are part of the F* group. Word-class errors are part of the G* group. For the purpose of tagging LINDSEI_CZ I added a separate tag for errors in the wrong choice of independent prepositions (LSP) as these appeared to be numerous.

Lexical errors form 33.28% of all errors identified in LINDSEI_CZ and are thus the second most populated group after grammatical errors. At least one lexical error was made by all of the speakers. Fig. 4–6 shows the lexical error counts for individual speakers, facilitating their comparison: 18 speakers made 1–5 errors, 20 speakers made 6–10, 8 speakers made 11–20, and 4 speakers made more than 20 errors.

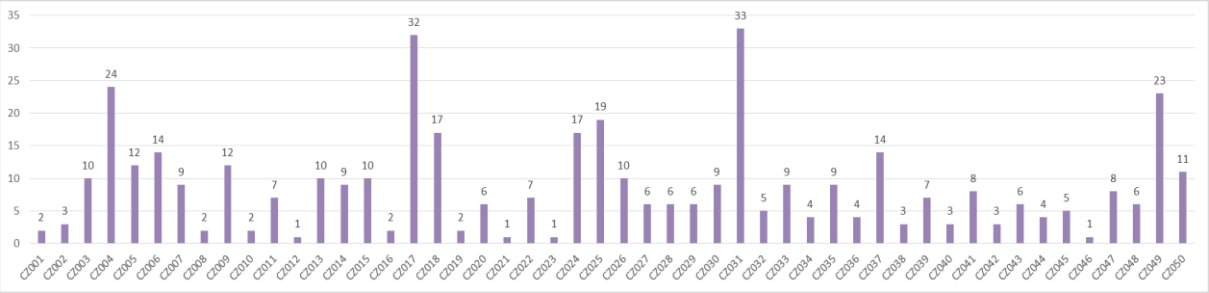

Fig. 4–6 Lexical error (L) counts per speaker. The numbers above the bars denote the number of lexical errors.*

Lexical errors form three subgroups — single word errors (LS*), phrasal errors (LP*) and connector errors (LC*). Their proportion is shown in Table 4–13.

Table 4–13 Lexical errors — absolute and relative frequencies (n = number of erring speakers)

Type	**Count**	**%**	**n**
Single-word errors (LS*)	290	66.8%	50
Phrasal errors (LP*)	131	30.2%	36
Connector errors (LC*)	13	3%	11
Total	**434**	100%	

There are more than twice as many lexical single errors than phrasal errors, whilst the number of lexical connector errors is negligible in comparison.

4.1.4.1 Lexical single errors (the LS* group)

The Louvain manual describes the lexical single errors as "conceptual, collocational or connotative errors in single words only" (Dagneaux et al. 2008, 31). They consist of three subgroups (see Table 4–14):

– lexical single errors proper (tagged (LS)) — *"the pronunciation is the main (erm) . (LS) setback $drawback$"*

– errors in the use of independent preposition (LSP) — *"it was (LSP) above $over$ seven hundred crowns"*

– false-friend errors (LSF) — *"he threw it out into one of the <starts laughing> (LSF) containers $dustbins$".*

Table 4–14 Single word errors — absolute and relative frequencies (n = number of erring speakers)

Type	Count	%	n
LS errors	144	49.7%	40
LSP errors	121	41.7%	42
LSF errors	25	8.6%	13
Total	**290**	**100%**	

Lexical single errors (LS) were committed by 80% of the speakers and constitute 48% of all single-word errors and 32% of all lexical errors in LINDSEI_CZ. 10 speakers made no LS errors, 30 speakers made 1–5 errors, 8 speakers made 6–10 errors, and 2 speakers made more than 10 errors.

Errors in the use of independent prepositions were committed by 84% of the speakers and constitute 43% of all single-word errors and 29% of all lexical errors in LINDSEI_CZ. 8 speakers made no LSP errors, 39 speakers made 1–5 errors, and 3 speakers made 6–10 errors.

False-friend errors were committed by 26% of the speakers and constitute 9% of all single-word errors and 6% of all lexical errors in LINDSEI_CZ. 37 speakers made no LSF errors, and the remaining 13 speakers made 1–5 LSF errors. Fig. 4–7 shows the distribution of LS* errors for all of the speakers.

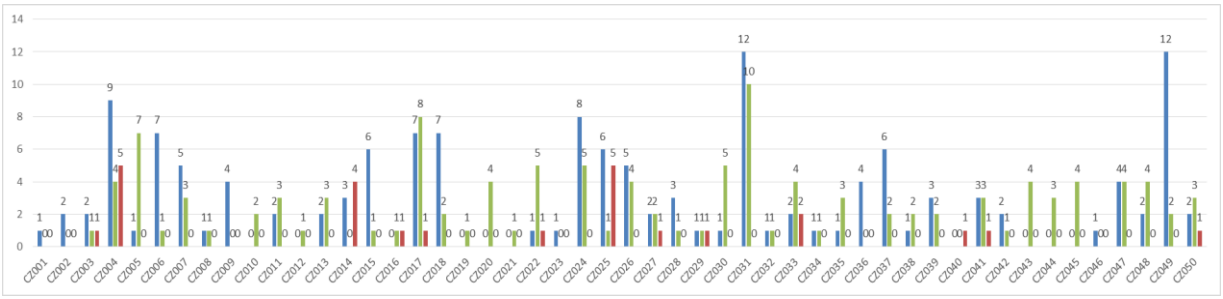

Fig. 4–7 Lexical single (LS) counts per speaker. The blue bars represent LS errors, the green bars LSP errors, and the red bars LSF errors.*

A closer look at the LS errors provides many opportunities for exploration and further categorization of these errors into subgroups. Four such subgroups were identified in my dataset: idiosyncratic examples (43.7%), confusing pairs (29.9), L1 transfer (21.5%), and the use of foreign words (4.9%).

The biggest group (43.7%) is made up of idiosyncratic examples which include unrepeated examples of lexical misuse whose source cannot be determined with absolute certainty. Problems are especially of collocational nature.

I can (LS) trace see the difference there

you go by (eh) by the underground (LS) take go (eh) . one one stop to

it was (er) (LS) maintained $organized$ very differently

There are a number of examples (29.9%) in which the error could be interpreted as a confusion with a similar word (the similarity might be either in form, the use, or the relation to an L1 expression). Such pairs are for example partially/partly, critics/critique, food/cuisine, setback/drawback, unsatisfied/dissatisfied, recorder/recording, impact/effect, talk/say and many other.

we (eh) got quite (LS) acquainted $familiar$ with each other

I . knew they were (LS) talking $saying$ (eh) something similar to me

The third group is formed by examples which can be explained as L1 transfer. A particular problem appears to be with semantically related concepts which are expressed by one lexeme in Czech but by different lexemes in English (e.g. the Czech ohodnotit/ocenit –> evaluate, appreciate,

prize, review, estimate, adjudicate...). Some of these pairs are well known to be commonly confusable (e.g. place/space, say/tell, opportunity/ocassion, learn/teach) and if a distinction were made here between mistakes and slips they would undoubtedly be classified as the latter.

The smallest group (4.9%) contains examples in which the speakers use a Czech word on failing to produce the required English expression. This strategy was undoubtedly helped by the fact the students knew both of the interviewers understood Czech.

I saw . (em) .. (LS) <foreign> sfinga </foreign> $the sphinx$ I'm not sure how to

The identification of prepositional errors is not unproblematic. Firstly, a distinction has to be made between dependent (these belong to the lexico-grammatical group of errors) and independent prepositions. Dependent prepositions typically introduce an object (*wait for somebody*) or a nominal complement (*an expert in early history*, *good at languages*). They form one semantic unit with the preceding word and their selection is not affected by the word they introduce. In this respect they differ from independent prepositions whose selection is affected by the relation (temporal, spatial or other) to the word they introduce (*at work, to work* etc.).

Secondly a decision has to be made whether the error is to be treated as an instance of wrong selection of a preposition or a wrongly used lexical phrase as in the expression **people in your age*. Only if the expression is idiomatic is it treated here as a phrase. Thus the expression *in the end* meaning *finally* is considered a phrase, whilst the locative expression *at the end* is not.

Of the total number of 121 errors in the use of independent prepositions 105 errors (87%) involved the prepositions *in*, *at* or *on.* These instances could be further broken down into the following groups:

– **on the picture/painting/drawing/portrait* (48 instances, 39.7% of the total LSP count);

– **in university/school* (19 instances, 15.7% of the total LSP count);

– various other instances (38 instances, 31.4% of the total LSP count).

The high incidence of the **on the picture* error is attributable to the design of the last task (a picture description). The error is made by 22 (44%) speakers and is probably a L1 transfer error (cf. the Czech *na obrázku*). There are further 23 errors which appear to be a result of a direct translation from Czech, e.g.:

– *took the ferry (LSP)* ***on*** *to the other side* (cf. **na** druhou stranu);

– *I could see it (LSP)* ***on*** *in the boy's behaviour* (cf. viděla jsem to **na** chlapcově chování);

– *what do people (LSP)* ***in*** *of your age do* (cf. co dělají lidé **ve** tvém věku);

– *that was really sad (LSP)* ***to*** *for me* (cf. the Czech dative bylo **mi** z toho smutno);

– *to do maturita (LSP)* ***from*** *in English* (cf. složit zkoušku **z** angličtiny).

The relatively high number of transfer errors would appear to justify the use of pedagogical materials in the classroom specifically designed to target these confusing areas.

Other prepositions apart from *in*, *at* and *on* are harder to classify. Some are the result of transfer (see the last two examples above), others are non-systematic. There does not appear to be another subgroup of typical errors. It is, however, somewhat surprising to see that even advanced learners can make errors involving prepositional use commonly taught at lower levels of proficiency (e.g. **on Christmas, *on December*, *except from Latin*).

Pearson correlation coefficient was calculated to establish whether there exists a correlation between the speaker's error rate and the number of prepositional errors. There was a strong positive correlation between the two variables ($r = .537$, $n = 50$, $p < .0005$), showing that less advanced learners in our corpus are prone to make more prepositional errors.

4.1.4.2 Lexical phrase errors (the LP* group)

When errors affect multiple words the Louvain manual marks them with the lexical phrase tag (LP). Lexical phrase errors are of the following types:

– errors involving the use of phrasal verbs — *"the conversation (LP) becomes about $turns to$ anything but";*

– a failure to use a set expression and replacing it with a non-idiomatic phrase — *"I really (LP) like . more $prefer$ the version";*

– an erroneous coinage of a lexical phrase which does not exist in English and the single semantic concept has to be expressed by paraphrasing it in English — *"and I'm (LP) studying master $doing my master's$";*

– an inaccurate reproduction of a set phrase — *"the lady is not really happy about how the picture (LP) looks like $looks$"*

Lexical phrase errors consist of two subgroups:

– lexical phrase errors proper (tagged (LP)) — "*so I (LP) made the exams $did the exams$"*

– lexical phrase false-friend errors (LPF) — "*that you are (LPF) on a good way $heading in the right direction$"*

The proportion of these groups is shown in Table 4–15.

Table 4–15 Lexical phrase errors — absolute and relative frequencies (n = number of erring speakers)

Type	Count	%	n
LP errors	125	95.4%	35
LPF errors	6	4.6%	4
Total	**131**	**100%**	

LP errors (LP) were committed by 70% of the speakers and constitute 95% of all phrase errors and 28% of all lexical errors in LINDSEI_CZ. 15 speakers made no LP errors, 28 speakers made 1–5 errors, 6 speakers made 6–10 errors, and 1 speaker made more than 10 errors.

Lexical phrase LPF errors were made by 8% of the speakers and constitute 1.4% of all phrase errors in LINDSEI_CZ.

4.1.4.3 Lexical connector errors (the LC* group)

Errors affecting the use of connectors are tagged as (LC) and divided into three subcategories: coordinating conjunctions (LCC), subordinating conjunctions (LCS) and logical connectors (LCL).

Only two of these errors are represented in LINDSEI_CZ:

– errors involving subordinating conjunctions (LCS) — "*it would be very convenient (LCS) when if he married"*

– errors involving a complex logical connector (LCLC) — "*because (LCLC) more I learn about it $the more I learn about it the more$ I see the processes of thinking*"

Table 4–16 Lexical connector errors — absolute and relative frequencies (n = number of erring speakers)

Type	Count	%	n
LCS errors	11	84.6%	10
LCLC errors	2	15.4%	2
Total	**13**	**100%**	

LCS errors were committed by 20% of the speakers and constitute 85% of all lexical connector errors and 13% of all lexical errors in LINDSEI_CZ. LCLC errors were comitted by 4% of the speakers and constitute 15% of all lexical connector errors and .5% of all lexical errors in LINDSEI_CZ.

4.1.5 Word redundant, word missing and word order errors (the W* group)

This group of errors consists of utterances containing redundant words (tagged WRS[31]), missing words (tagged WM) and word-order errors (tagged WO). The frequencies are seen in Fig. 4–17.

Table 4–17 The W group errors — absolute and relative frequencies (n = number of erring speakers)*

Type	Count	%	n
WRS	4	8%	3
WM	11	22%	9
WO	35	70%	19
Total	**50**	**100%**	

It is essential to point out that in accordance with the Louvain error-tagging manual word-order errors in LINDSEI_CZ are split into two groups: the WO-group discussed above, and the GADVO-group which includes all instances of wrong word order involving adverbs. If errors in these two groups were counted together (as would appear somewhat more logical), word-order errors would form 5.6% of the total error count and thus be more frequent than errors in the lexico-grammatical group.

[31]The manual further distinguishes between single redundant words (tagged WRS) and multiple redundant words (tagged WRM). The latter did not appear in LINDSEI_CZ.

An attempt to categorize word-order errors into subgroups revealed that there is one recurrent type of error, namely inverted word order in embedded questions (46% of all word-order errors), which can be considered intralingual as it is probably committed as a result of a false analogy with the standard question word order and triggered by the use of the interrogative pronoun:

I don't even know where (WO) am I $I am$ going to be living

The other errors are less frequent and of diverse nature. Most are probably slips of the tongue and only result from planning pressure. There are 4 instances[32] of the unnatural sounding combination of word-order infelicity and the tautological use of the first person plural *we* instead of the singular *I*. This is undoubtedly a case of L1 transfer and was labelled as Czenglish by Sparling (1991).

(GPP) (WO) we were there with my sister $my sister and I were there$ (cf. the Czech *byli jsme tam se sestrou*)

4.1.6 Infelicities (the Z group)

The Louvain error-tagging manual distinguishes between true errors and cases of infelicity which are described as "long(ish) [...] chunks of text" which are not really erroneous but sound so unnatural, unidiomatic, clumsy or foreign that a native speaker would feel the need to reformulate them (p. 42). At the same time it is not possible to assign the problem to one particular error tag. The identification of infelicities is subjective. The errors are mostly of lexical nature but do not fit the criteria for inclusion in the LS or LP groups. Every effort was made in this thesis to reduce the number of inclusions in this group to a minimum to avoid the risk of making the group "an ill-defined dustbin category" (ibid., 43). The result is, however, far from satisfactory and serves as a good example of the difficulties involved in identifying errors and especially border-line cases. Only a more detailed error-tagging system would enable a further reduction of the number of examples included in the group.

[32]I tagged these as WO as the use of the pronoun seems to be triggered by the word order. If the speaker started with the other part of the subject he would not use *we*, e.g ?*my sister and we*).

All in all, there are 30 inclusions (committed by 16 speakers) in the Z group. Owing to their diverse nature they cannot easily be separated into smaller groups. They are mostly examples of unidiomatic utterances (see the first three examples below), problems with logic (the fourth example) and L1 interference examples (see the last three examples below).

we plan . going there again . (erm) (Z) at the beginning of the next year $early next year$

(DTG) (Z) it was like on = (GA) 0 a one-year . old building $the building was about a year old$

these two books are . my (DTG) (GADJCS)(Z) most favourite ones $absolute favourites$

a man . who . (Z) could be a possible husband $is possibly her husband"

so he had (Z) he had the only idea $the only thing that occured to him"

where Dustin Hoffmann comes and (Z) jumps from the window to the church $jumps in through the church window$

I think (Z) this is quite clever to say $this is a clever thing to say$

4.2 Error rate — results

The fifty speakers in LINDSEI_CZ committed a total of 1,299 errors. The absolute frequency of errors ranges between 5 errors (speaker CZ023) and 89 errors (speaker CZ017) showing the

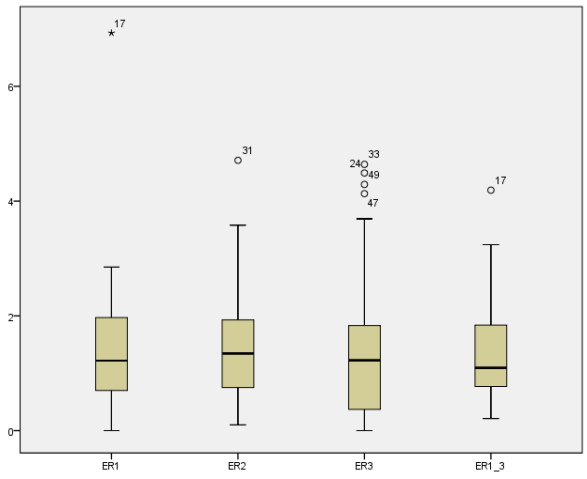

Fig. 4–8 Boxplots showing the dispersion in the error rates in Tasks 1, 2 and 3, and in the global error rate (T1–3)

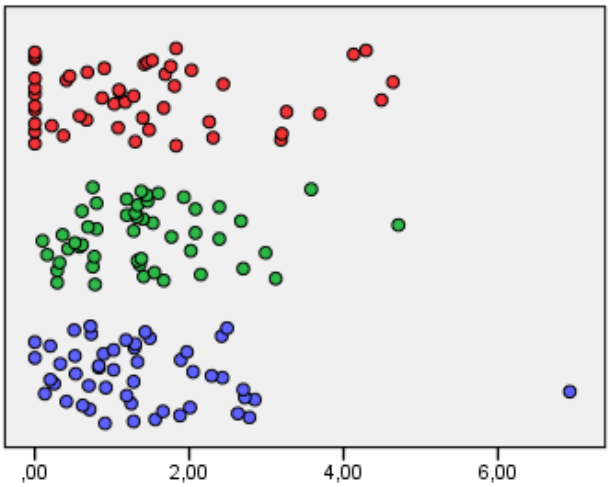

Fig. 4–9 Scatterplots showing the overlap of the error rates in Tasks 1 (marked blue), 2 (green) and 3 (red)

speakers greatly vary in the numbers of errors they produce. As the length of the interviews and the proportions of the individual tasks are not constant, I will henceforth work with a normalized

frequency of errors per hundred words (phw), i.e. the so-called error rate (ER). Table 4–18 and the boxplots and scatterplots in Figs. 4–8 and 4–9 show the distribution and dispersion of the error rate values.

Table 4–18 Error rates in Tasks 1, 2 and 3, and global error rates (T1–3)

Task	T1	T2	T3	T1–3
Minimum	0.00	0.10	0.00	0.21
Maximum	6.93	4.71	4.64	4.19
Mean	1.39	1.43	1.42	1.39
SD	1.14	0.94	1.31	0.87
Median	1.22	1.35	1.22	1.10

Whilst the results of the error analysis (see section 4.1) show that task variability has an effect on the type of errors the learners commit (in Task 3, unlike in Tasks 1 and 2, the learners produced more lexical than grammatical errors) the diagrams in Figs. 4–8 and 4–9 are considerably similar, suggesting that error rate is not affected by task variability. To prove this hypothesis I carried out a one-way repeated measures ANOVA which, indeed, proved no significant effect for task variability (Wilk's lambda = .998, F (2,48) = .038, p = .962). In the subsequent analyses I will therefore work with the global rate for each speaker (i.e. ER1_3), and only use the task error rates where individual tasks are analyzed and compared. This can also be illustrated by Table 4–19 which is the result of a comparison of ERs for each speaker's performance in the individual tasks and provides information about the observed difficulty of the tasks. It shows that 22 speakers had the highest ER in Task 3 but that at the same time for 18 different speakers Task 3 was the least problematic.

Table 4–19 Speakers' accuracy in the three tasks (the numbers provide counts of speakers who had the most, medium and least accurate performance in Tasks 1–3)

	Hardest	Medium	Easiest
Task 1	14	18	18
Task 2	14	22	14
Task 3	22	10	18

4.2.1 Accuracy bands

The large dispersion in the error rate shows that as regards accuracy our group of learners is fairly heterogeneous. 5 speakers committed fewer than 10 errors, 19 speakers 11–20 errors, 7 speakers

21–30 errors, 4 speakers 41–50 errors, and 4 speakers more than 50 errors. Clearly, some of our advanced learners are more advanced than others, and some should possibly even not be considered advanced. This is not an unusual finding. Carlsen (2012) speaks of proficiency in learner corpora as of a fuzzy variable. Similar claims are made e.g. by Thomas (1994), Callies (2009), and Götz (2015). Whilst we could consider this a weakness in the design of our corpus, such a dispersion is a highly realistic picture of typical language classes which are rarely homogeneous. This seeming weakness also gives us the opportunity to compare the differences between our speakers and investigate whether the weakest speakers in our dataset share any traits with their more advanced colleagues.

To this end I divided the 50 speakers' ERs into five accuracy bands which were defined by the numbers of standard deviations subtracted from (for those with an ER smaller than the mean) or added (for those with an ER greater than the mean) to the mean ER (see Table 4–20).

Table 4–20 Accuracy bands defined by ER ranges

Accuracy band	min. ER	max. ER	Distance from the mean	n
L1	.00	.53	0 to -1SD	7
L2	.53	1.39	-1SD to mean	22
L3	1.39	2.26	mean to +1SD	14
L4	2.26	3.13	+1SD to +2SD	5
L5	3.13	4.10	> +2SD	2

The first band (L1), with the most accurate speakers, is defined by the range of ER between zero and .53 errors (-1SD) phw. It has 7 speakers. The second most accurate group (L2) commits between .53 (–1SD) and 1.30 (the mean) errors phw. It has 22 speakers. The third group (L3) is in the range of 1.39 (the mean) to 2.26 (+1SD) errors phw. It has 14 speakers. The fourth group commits between 2.26 (+1SD) and 3.13 (+2SD) errors phw. It has 5 speakers. The least accurate group's ER is above 3.13 (+2SD) errors phw and has 2 speakers.

4.2.2 Persistent errors

The division of speakers into accuracy bands described in the previous section enables comparisons of these groups of students and the identification of the most problematic areas. To achieve this goal, I calculated the numbers of errors for each error category within each accuracy

band. The raw and the normalized frequencies are shown in Table 4–21. The percentages are calculated as a ratio of the number of errors within each error category divided by the total error count for the respective accuracy bands.

Table 4–21 The frequency of the main error categories within accuracy bands. (L1–L5 — most to least accurate groups of students)

	F		G		X		L		W		Z		Error count
L1	0	0%	35	57.4%	6	9.8%	12	19.7%	6	9.8%	2	3.3%	**61**
L2	1	.3%	216	55.5%	20	5.1%	127	32.6%	18	4.6%	7	1.8%	**389**
L3	2	.5%	253	57.4%	24	5.4%	140	31.7%	14	3.2%	8	1.8%	**441**
L4	0	0%	118	48.2%	15	6.1%	100	40.8%	4	1.6%	8	3.3%	**245**
L5	0	0%	88	54.0%	6	3.7%	55	33.7%	9	5.5%	5	3.1%	**163**
Totals	**3**		**710**		**71**		**434**		**51**		**30**		**1299**

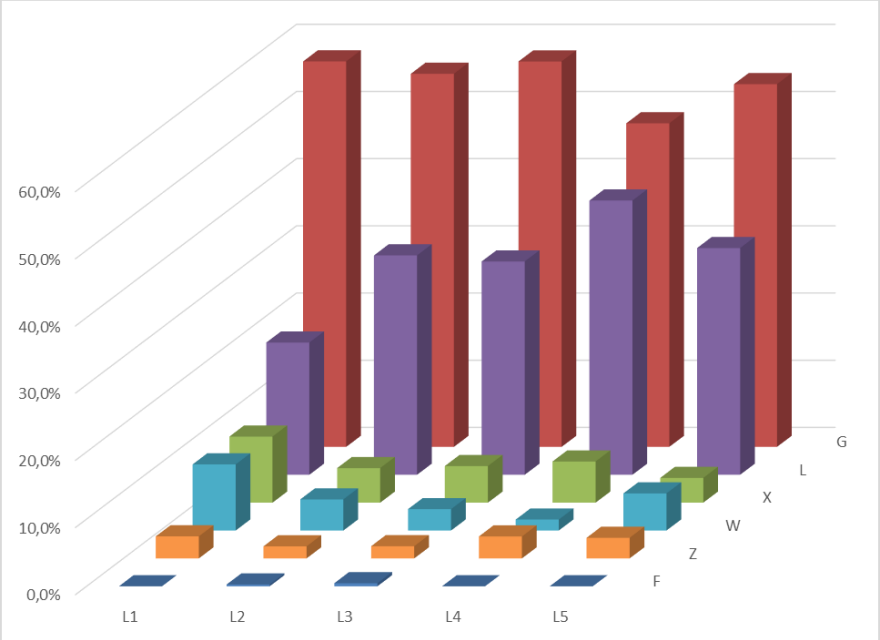

Fig. 4–10 The distribution of the main error categories within accuracy bands

As the results of the error analysis described in sections 4.1, 4.1.2 and 4.1.4 revealed, the most frequent errors in our corpus are grammatical and lexical. Fig. 4–10, which is a visualization of the relative frequencies provided in Table 4–21, shows that this applies to all of the students from the least (L5) to the most accurate (L1) ones. Whilst the results would seem to suggest that the most accurate group has a lower frequency of lexical errors, such a result might be affected by the overall low count of the errors in this accuracy band.

The same method can be used for the identification of persistent errors and help us determine whether the overall finding that errors in the use of articles and tenses are the most frequent ones applies in all of the accuracy bands. To this end I calculated the numbers of errors for each grammatical error subcategory (i.e. each G* tag) and within each accuracy band. As the resulting graph would be too wide to display on a page, I subsumed some of the related subcategories into one. Thus tense errors include both GVT and GVTA (tense agreement), and all determiner, pronominal, adverbial, nominal, conditional, and adjectival subgroups are subsumed into their own respective groups. The results are shown in Fig. 4–11.

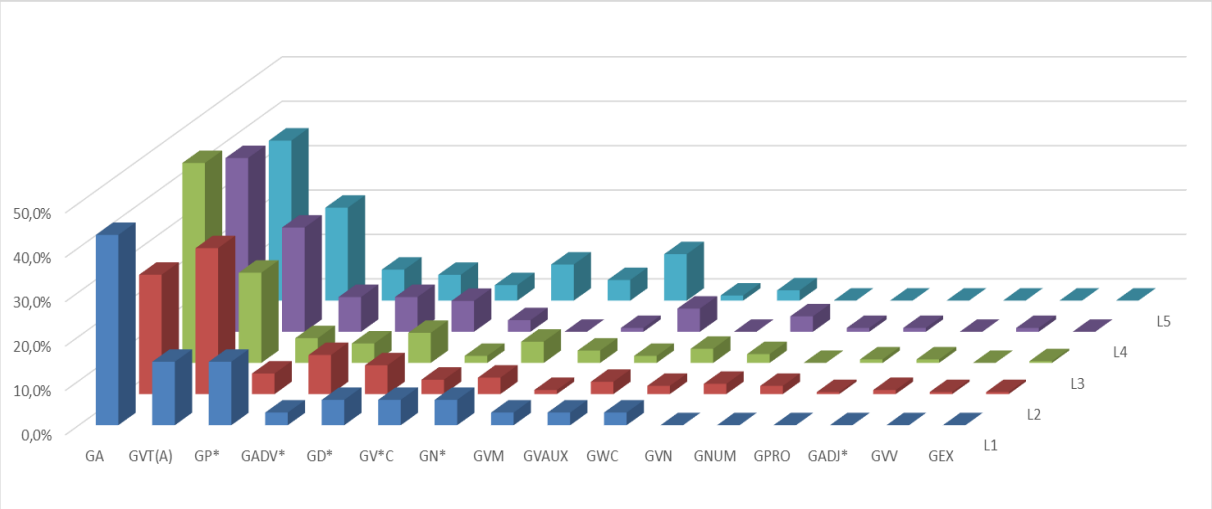

Fig. 4–11 The distribution of grammar error types within accuracy bands

Fig. 4–11 rather clearly shows that problems with article errors affect all of the groups and article errors may thus be considered persistent. The same applies to verb tense errors, although there appears to be a decrease in their frequency in the most advanced group. The learners regularly commit errors in the use of pronouns, adverbs, determiners and the conditional. The remaining types of errors are less numerous which means they are committed by smaller numbers of speakers and with a smaller frequency. Consequently no generalisations ought to be made.

4.2.3 Accuracy correlations

All of the most frequently occurring error types were found to correlate with the error rate. Large positive correlations were found between ER and the frequency of article errors (r = .728,

$p < .0005$), lexical single errors ($r = .710$, $p < .0005$), lexical phrase errors ($r = .695$, $p < .0005$), and verb tense errors ($r = .728$, $p < .01$). Such result is, however, not surprising given the fact that these errors present by far the most frequent error types in the dataset.

4.3 Accuracy conclusion

The identification of 1,299 errors in LINDSEI_CZ provided an opportunity to perform quantitative and qualitative analyses of these errors and answer the research questions and hypotheses formulated as a result of the pilot study.

The group of 50 students examined in this thesis proved to be rather heterogeneous with the most accurate one producing as few as one error every 476 words and the least accurate one as many as one error every 24 words. Such a dispersion raises important questions about the design of learner corpora and especially the need to introduce more rigorous ways of establishing and guaranteeing the participants' proficiency (cf. e.g. Carlsen 2012). At the same time the sample of our data is closer to classroom reality and is the true representation of the varying standards of the students of English philology, who were the target of my research.

The results of my error analysis are very similar to those in the German subcorpus of LINDSEI (henceforth LINDSEI_GE). Table 4–22 shows the comparison of the two subcorpora.

Table 4–22 Comparison of selected error frequencies in LINDSEI_GE and LINDSEI_CZ

Global error types					**Lexical**					**Grammatical**				
	GE		**CZ**			**GE**		**CZ**			**GE**		**CZ**	
	Count	**%**	**Count**	**%**		**Count**	**%**	**Count**	**%**		**Count**	**%**	**Count**	**%**
F	24	2%	3	0%	**LS**	130	27%	144	33%	**GA**	138	22%	265	37%
G	627	47%	710	55%	**LSP**	114	24%	121	28%	**GVT***	233	37%	177	25%
X	67	5%	71	5%	**LSF**	51	11%	25	6%	**GP***	52	8%	44	6%
L	480	36%	434	33%	**LP**	91	19%	125	29%	**GADV***	24	4%	52	7%
W	114	9%	51	4%	**LPF**	90	19%	6	1%	**GD***	24	4%	44	6%
Z	23	2%	30	2%	**Other**	4	1%	13	3%	**Other**	156	25%	128	18%
Total	**1335**		**1299**		**Total L***	**480**		**434**		**Total G***	**627**		**710**	

Despite claims (e.g. Lennon 2000, Agustín Llach 2011) that lexical errors[33] are generally more numerous than grammatical ones, in LINDSEI_CZ (and LINDSEI_GE) it is grammar errors that occur with a higher frequency. Here, the most frequent problem is the use of articles, which confirms Dušková's (1969) findings. It is especially article omission which appears to be the frequent cause of errors, which suggests that this is an area affected by L1 transfer, especially if we compare our findings with the study of LINDSEI_GE (Götz 2015) which finds the use of articles much less problematic (GA errors form 22% of LINDSEI_GE grammar errors and 37% of LINDSEI_CZ grammar errors). We might, however, also speculate that articles are taught better at German schools.

The second most frequently appearing grammar error is the use of verb tenses. Here it is especially the use of the present perfect. Whilst this might be partly affected by the design of the tasks in which the observed learners participated, this is a notoriously problematic area for Czech students (cf. Götz (2015) who finds a smaller frequency of present perfect errors). Whilst some of these errors may be explained as L1 transfer, there are also several examples of adverbials triggering off the use of the present perfect and indicating that an undue focus in teaching materials on the role of adverbials in the present perfect might be the source of at least some part of these problems (cf. Granger 1999).

Other types of grammatical errors are committed by smaller numbers of students, which suggests that these are non-systematic errors.

In the 71 lexico-grammatical errors verb complementation proved to be the most problematic, but the errors affected only about 50% of the speakers. It is worth pointing out that this is a border-line category and the question arises whether these errors ought not to be subsumed within the category of grammar errors instead.

Lexical errors are the second most frequent but their nature confirms Dušková's (1969, 24) statement that "errors in lexis presented a much less homogeneous material for study than errors in

[33]Care must be taken in interpreting such statements unless we take a detailed look at how errors and lexical errors are defined. Many studies reported by Agustín Llach (2011) include in the category of lexical errors such errors that would be classified as grammatical according to the Louvain error-tagging manual.

grammar". Consequently, it is much harder to design a satisfactory classification system, especially as the Louvain manual only lists a handful of very broad categories. Thus, amongst the 434 lexical errors identified in LINDSEI_CZ, we can clearly see that the majority are formed by lexical single errors, amongst which errors in the use of prepositions play an important role. More detailed analysis reveals that whilst there is a number of idiosyncratic, hard-to-classify errors, there also exists a large group of errors in the use of words which are easy to confuse (especially because of their similarity to other words, both in L1 and L2). This is an area which needs a much deeper analysis based on a sound system of classification. The results are comparable with LINDSEI_GE.

The analysis of error rates did not reveal significant differences in the error rates in the different tasks, which confirms Salaberry and Lopez-Ortega's (1998) findings that more advanced groups are less affected by different task designs even though different types of language are produced (cf. Tracy-Ventura and Myles 2015).

The division of the speakers into accuracy bands according to their overall error rate provided an opportunity to prove that the error types identified as most frequent in the error analysis are the most problematic types for all students from the least to the most advanced ones. Thus, even the most advanced students regularly commit errors especially in the use of articles and tenses. These errors can therefore be considered not only systematic but also persistent, proving resilient to improvement even after many years of study. Owing to their high frequency and their presence across the board they also have a predictive power for the overall error rate.

To conclude, let us summarize the answers to the research questions (RQs) formulated in section 3.1.2. In answer to RQ1, whether persistent errors may be identified, the study showed that this was especially the use of grammar and lexis (see sections 4.1 and 4.2.1). As regards grammar, the use of articles and tenses proved to be the most difficult (see 4.1.2), as for lexis, lexical single errors and prepositional errors were the most frequent (see 4.1.4). In answer to RQ2, whether problematic areas were linked to particular tasks, the study showed (see 4.1) that lexical errors were more frequent than grammatical in Task 3, elsewhere and globally the opposite proved to be the case. As regards RQ3 and the questions regarding the homogeneity of the group of students, the group's dispersion,

as evidenced by the error rate, proved to be large (see 4.2) in all of the tasks. As for RQ4 regarding the link between the level of accuracy and particular types of errors, lower-level students proved to commit more prepositional errors (see 4.1.4.1) but it could not be proved that in our sample there existed such errors which were only committed by less accurate students. In answer to RQ5 as to the correlation between error rate and particular type of errors, higher error rate did not appear to have a predictive power as the most frequent errors proved to be present in all of the accuracy bands. As for RQ6 and the possibility of establishing the sources of errors, these could not be ascertained with certainty, but the analyses showed that L1 transfer could be tentatively linked to article omission (see 4.1.2.2), noun number (see 4.1.2.3), dependent prepositions (see 4.1.3), single lexical items (see 4.1.4.1), prepositions (see 4.1.4.1), and some of the examples of infelicities (see 4.1.6). Word order errors with embedded questions may be intralingual (see 4.1.5) but they could also be teaching-induced.

The results of these analyses prove the following hypotheses formulated in section 3.1.2:

1. Grammatical errors are the most frequent error type. (see 4.1)
2. Lexical errors are the second most frequent error type. (see 4.2)
3. Errors in the use of articles are the most frequent type of grammatical errors. (see 4.1.2)
4. Errors in the use of tenses are the second most frequent type of grammatical errors. (see 4.1.2)
6. Errors in the use of prepositions are the second most frequent type of lexical errors. (see 4.1.4)
7. There is a correlation between errors in the use of articles and error rate. (see 4.2.3)
8. There is a correlation between lexical errors and error rate. (see 4.2.3)
9. There is a correlation between errors in the use of lexical phrases and error rate. (see 4.2.3)
10. There is a correlation between errors in the use of tenses and error rate. (see 4.2.3)

The following hypotheses formulated in section 3.1.2 were not proved:

5. Errors in the use of lexical phrases are the most frequent type of lexical errors. (see 4.1, lexical single errors are more frequent)
11. Task one has the strongest effect on accuracy. (see 4.2, no significant effect was found)
12. Task three has the weakest effect on accuracy. (see 4.2, no significant effect was found)

5. Fluency — results and analyses

The following section presents the results of fluency measurements in LINDSEI_CZ and in LOCNEC thus comparing non-native and native speech performance. As shown in the pilot study, fluency is operationalized as speed and breakdown fluency and a small selection of some of the most salient features to be measured is made, namely speech rate, frequency of filled and unfilled pauses and the mean length of runs.

5.1 Fluency — speech rates — results

Using the techniques described in the pilot study, the speech rate for unpruned words (i.e. all words uttered including filled pauses) was measured in words per minute (wpm). Non-verbal sounds were excluded as were long periods of silence (e.g. the time the students spent looking at the drawings in Task 3). I will firstly present a comparison of the variance in task performance, and then provide bar charts and descriptive statistical figures to present the results of the measurements in the separate tasks both for LINDSEI_CZ and LOCNEC.

The results of the pilot study showed that a difference in speech rate is to be expected between the three different tasks. The variances in the non-native-speaker and native-speaker speech rates are shown in the boxplots in Fig. 5–1.

To compare the speech rate in the three tasks performed by the non-native speakers a one-way repeated measures ANOVA was carried out. A significant effect was found for task variability, Wilk's Lambda = .401, F (2,48) = 35.8, $p < .0005$. A very large effect size is evidenced by the multivariate partial eta squared of .6. To determine which of the differences is statistically significant all of the tasks were compared in a paired-samples t-test

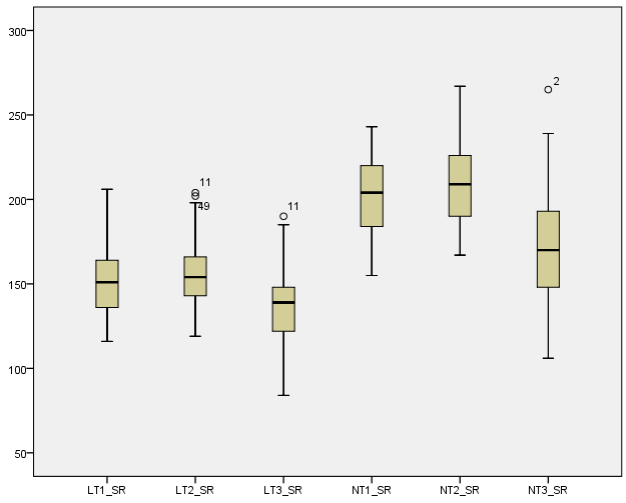

Fig. 5–1 Boxplots showing non-native (L) and native (N*) speech rate for each task separately (*T1—3). The y axis marks speech rate in words per minute (wpm)*

(2-tailed). A statistically significant difference was found between all tasks. For Tasks 1 and 2, t (49) = -3.38, $p < .05$; for Tasks 1 and 3, t (49) = 6.55, $p < .0005$; and for Tasks 2 and 3, t (49) = 6.5, $p < .0005$ with a moderate effect size for Tasks 1 and 2 (partial eta squared of .1), and a large effect size for the other two comparisons (partial eta squared of .3)

A comparison of the three tasks performed by the native speakers using the same statistical tests proved a significant effect for task variability, Wilk's Lambda = .363, F (2,47) = 41.28 with a very large effect size (multivariate partial eta squared = .64). The paired samples t-tests revealed significant differences between all tasks. For Tasks 1 and 2, t (48) = -2.72, $p < .01$; for Tasks 1 and 3, t (48) = 6.49, $p < .0005$; and for Tasks 2 and 3, t (48) = 9.06, $p < .0005$ with a large effect size expressed by the partial eta squared of .07, .3, .46 respectively.

5.1.1. Fluency — non-native-speaker speech rates — results

As the variance tests showed significant difference in speech rate in the different tasks, all of the tasks are treated separately. Figures 5–2—5–4 show bar charts illustrating the non-native speakers' performance in the individual tasks, the bars are ordered from the slowest speech rate to the fastest.

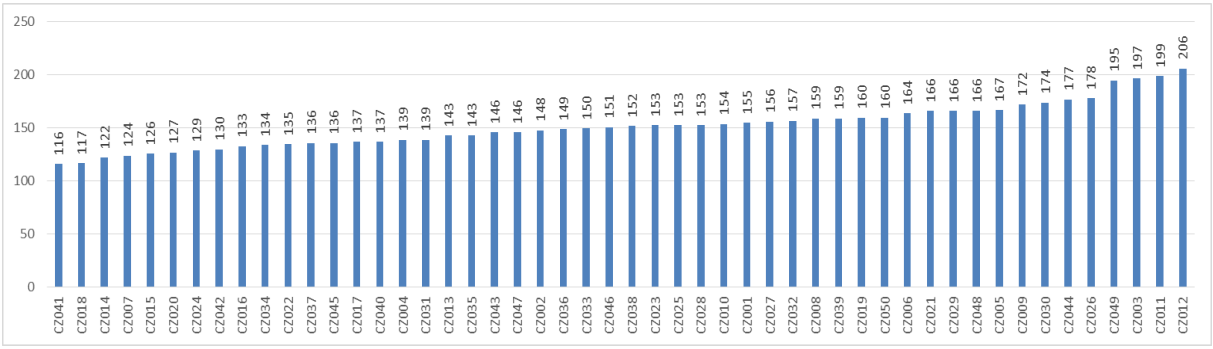

Fig. 5–2 Non-native speech rates in Task 1 for all LINDSEI_CZ speakers. The figures above the bars represent SR (wpm)

The speech rates for Task 1 for non-native speakers range between 116 and 206 wpm. The mean speech rate is 152 (SD = 20.97), the median is 151 wpm.

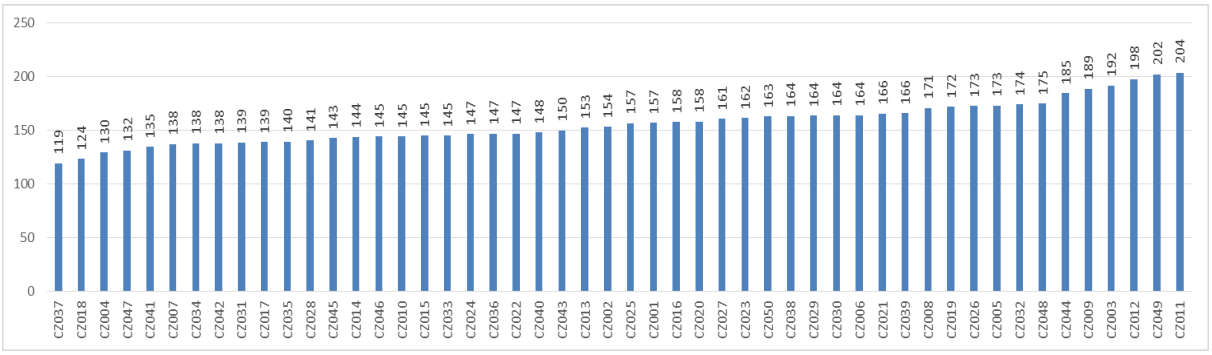

Fig. 5–3 Non-native speech rates in Task 2 for all LINDSEI_CZ speakers. The figures above the bars represent SR (wpm)

The speech rates for Task 2 for non-native speakers range between 119 and 204 wpm. The mean speech rate is 157 (SD = 19.72), the median is 155 wpm.

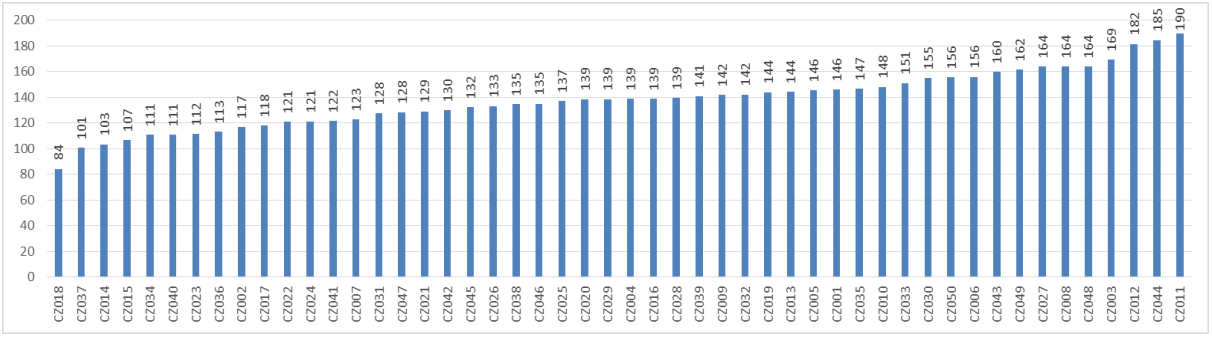

Fig. 5–4 Non-native speech rates in Task 3 for all LINDSEI_CZ speakers. The figures above the bars represent SR (wpm)

The speech rates for Task 3 for non-native speakers range between 84 and 190 wpm. The mean speech rate is 138 (SD = 22.09), the median is 139 wpm.

5.1.2 Fluency — native-speaker speech rates — results

For comparison, Figures 5–5—5–7 show bar charts illustrating the native speakers' performance in the individual tasks, the bars are ordered from the slowest speech rate to the fastest.

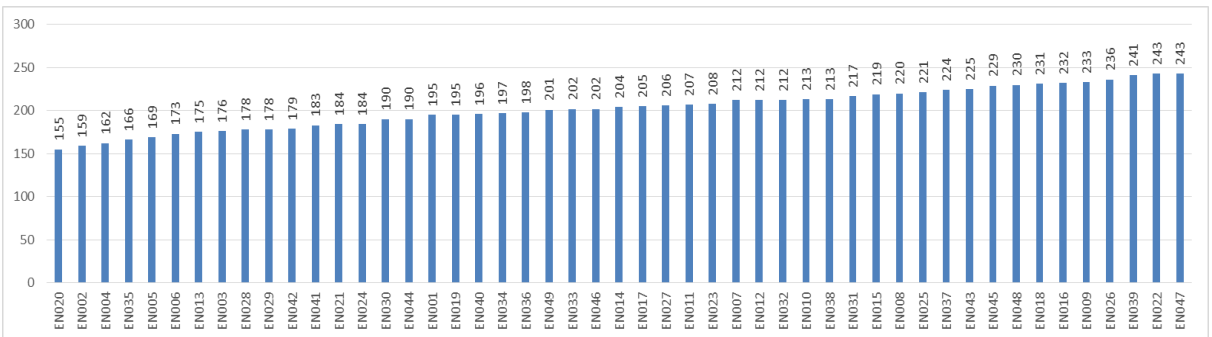

Fig. 5–5 Native speech rates in Task 1 for all LOCNEC speakers. The figures above the bars represent SR (wpm).

The speech rates for Task 1 for native speakers range between 155 and 243 wpm. The mean speech rate is 203 (SD = 23.51), the median is 204 wpm.

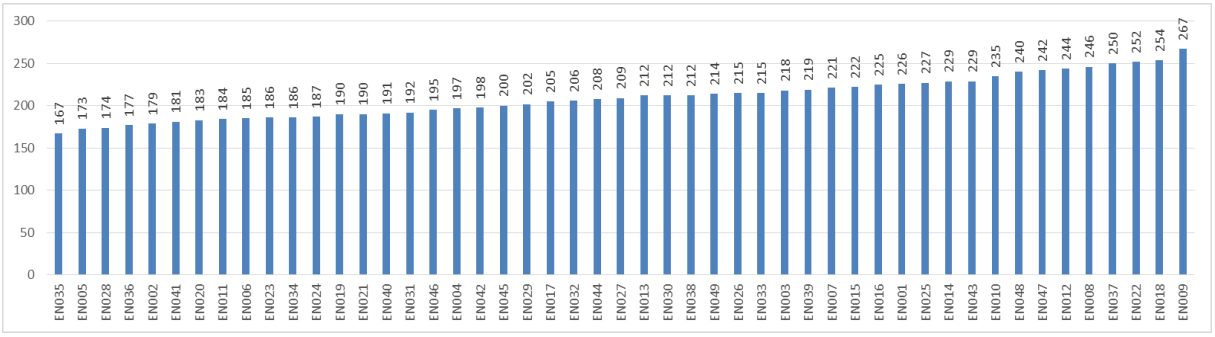

Fig. 5–6 Native speech rates in Task 2 for all LOCNEC speakers. The figures above the bars represent SR (wpm).

The speech rates for Task 2 for native speakers range between 167 and 267 wpm. The mean speech rate is 210 (SD = 24.53), the median is 209 wpm.

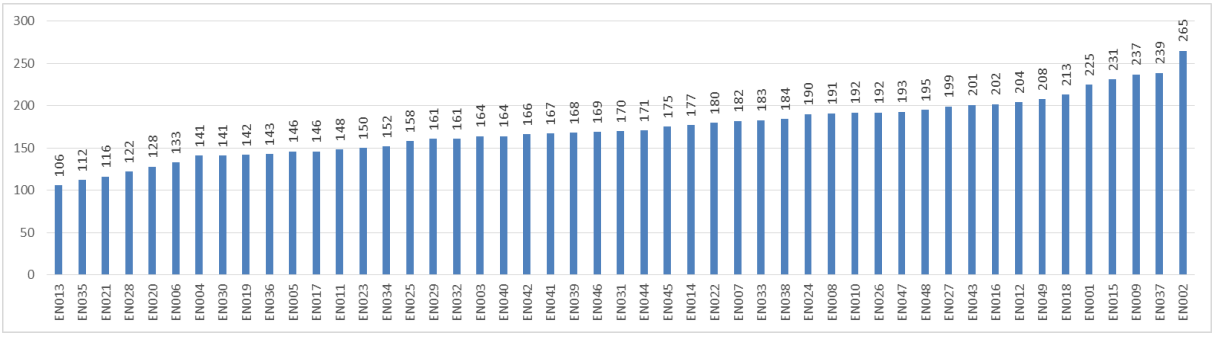

Fig. 5–7 Native speech rates in Task 3 for all LOCNEC speakers. The figures above the bars represent SR (wpm)

The speech rates for Task 3 for native speakers range between 106 and 265 wpm. The mean speech rate is 174 (SD = 34.49), the median is 170 wpm.

5.1.3 Fluency — speech rate — analysis

The comparison of native and non-native speech rates reveals that native speakers produce faster speech in all tasks. The mean speech rate for Task 1 in LINDSEI_CZ is 152 and for LOCNEC 203 wpm. For Task 2 it is 157 and 209 wpm respectively, and in Task 3 it is 138 and 173 wpm respectively. The results of t-tests prove that the differences are statistically significant ($p < .0005$). For Task 1 they are $t(48) = -11.18$, for Task 2 $t(48) = -12.1$, and for Task 3 $t(48) = -5.77$ with a large effect size (partial eta squared of .56, .6, and .26). We can therefore confirm hypothesis 1 formulated

above that the learners produce fewer words per minute in all tasks than the native speakers. These results are similar to Götz's (2013) comparison of LOCNEC and the German subcorpus of LINDSEI.

But is the learners' lower speech rate to be viewed as a problem as Götz (2013, 94) does? A more detailed exploration of our measurements offers an answer. If we explore the boxplots and scatterplots in Figs. 5–8 and 5–9 we will notice that owing to the wide dispersion of both of the samples of data there is an overlap between them. In Task 1, 20 fastest learners overlap with 20 slowest natives, in Task 2 this is true of 13 speakers, and in Task 3, 47 fastest learners overlap with 33 natives. Clearly, amongst the natives there are fast and slow speakers. Some of the slower natives speak more slowly than some of the faster learners. Still, a large proportion of the learners fail to speak as fast as even their slowest native counterparts. It would appear that speech rate could be used as a gauge for oral proficiency, and if our native speakers' lowest SRs start at 155 wpm, this could perhaps be the target value for spoken oral performance at the advanced level. In LINDSEI_CZ 20 learners (40%) have reached the target in Task 1, and 24 (48%) in Task 2.

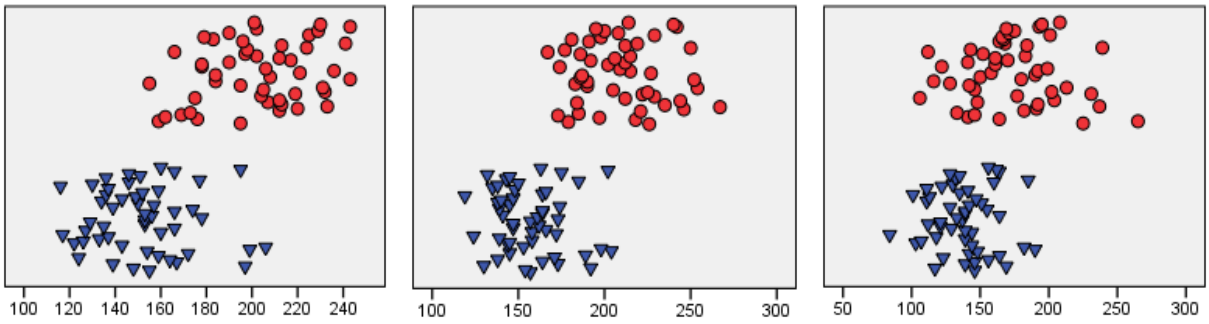

Fig. 5–8 Scatterplots showing the overlap between non-native (the blue triangles) and native (the red circles) speech rates in the three tasks. The X axis represents the speech rate (in wpm)

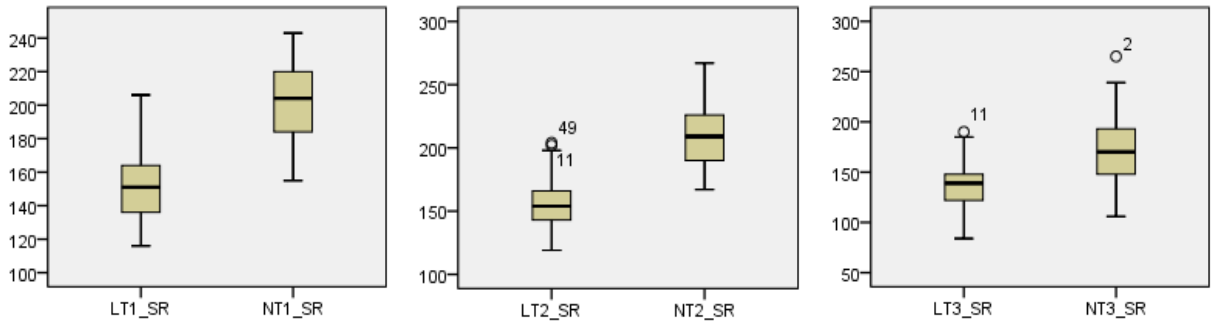

Fig. 5–9 Boxplots showing the comparisons of the dispersions of speech rates for Tasks 1—3 in LINDSEI_CZ and LOCNEC

The question arises as to what makes the learners' speech slower. It is tempting to say that it is the larger number of UPs. To test this hypothesis we would have to extract all silent sections and measure up the rest. This is, however, an extremely time-consuming venture made further complicated by the fact that pause boundaries are often unclear (e.g. the phonation of certain sounds starts before any sound is detectable). Accurate measuring of all pauses is thus practically impossible. But if we take into account the vast overuse of UPs by the learners, it is safe to assume that at least some part of the difference in the SR can be explained in this way. To test this hypothesis and determine the size of the correlation I investigated the relationship between SR and UP rate using the Pearson product-moment correlation coefficient. The results are shown in Table 5–1.

Table 5–1 Correlations in tasks between SR and UP rate

	Pearson r	Strength of correlation	p	Coefficient of determination
LT1	-.489	Medium	< .05	24%
LT2	-.425	Medium	< .05	18%
LT3	-.636	Large	< .05	40%

	Pearson r	Strength of correlation	p	Coefficient of determination
NT1	-.367	Medium	< .05	13%
NT2	-.249	N/A	> .05	N/A
NT3	-.558	Large	< .05	31%

The table shows that there is at least a medium-strength negative correlation between UP rate and SR, which means that faster speakers use fewer UPs. These results are statistically significant ($p < .05$) in all cases except in Task 2 for native speakers ($p > .05$). As is apparent from the coefficient of determination (last column of the table) the UP rate in our tasks can explain between 13 and 40% of the variance in speech rate. Thus, it is true that the learners speak more slowly partly because they produce more pauses. These results are also clearly visible in Fig. 5.10.

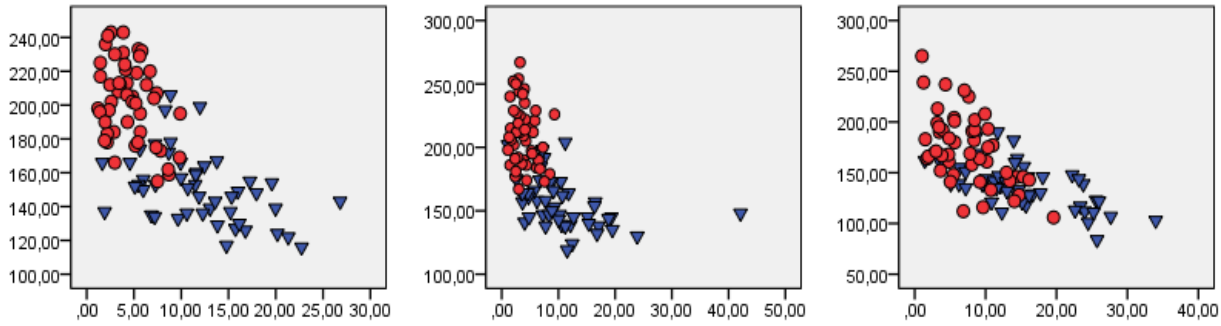

Fig. 5–10 Scatterplots showing the correlation between UP rate and speech rate

As no other variables measured in our study could explain the other parts of the variance in SR I carried out a perceptive test to determine other possible factors. Whilst this is only based on my subjective impression, I cannot refrain from observing that the native speakers appear to articulate with somewhat greater lightness and ease. Such as can be observed with the fastest learners. Interestingly, the perceptive test revealed that the fastest speakers amongst the learners are also those whose accent I would label as the most native-like. Such evaluations are, however, unreliable and difficult to operationalize. If, however, such a hypothesis were proved we could assume that pronunciation practice (perhaps including reading out loud) might lead to gains in speed.

As illustrated by the t-test comparisons in section 5.1, task variability plays a role in the performance of both the natives and the learners. The small degree of variation between Tasks 1 (designed as a monologue) and 2 (a dialogue) can be attributed to the fact that for a number of reasons (e.g. the students' reticence) in Task 1 the interviewer often interfered by asking questions and thus turned the task into more of a dialogue. In LOCNEC, the interviewer was naturally more active and possibly failed to see the essentially monological design of Task 1 as its goal. Thus many of the interviews in Task 1 are rather dialogues. Clearer instructions to both the interviewers and the interviewees would possibly have resulted in more obvious distinctions between the two tasks. Task 3 resulted in the slowest SR for both groups, which shows that when speakers have to analyse external factors (such as working out what is happening in a picture story), speech may be slowed down.

The speech rate noticeably varies not only from task to task but also within them. Speaking is somewhat like driving a car on a country lane. Longer stretches of straight road allow the driver to put the foot down, whilst obstacles in the form of bends or other vehicles necessitate more careful progress. The resulting average speed might be far from what would actually be possible in ideal conditions. Speech is just like that. Easier chunks of language such as standard responses to questions and preassembled formulas roll off the tongue automatically and with great speed whilst more complex thoughts requiring greater linguistic sophistication call for more time. The parallel with driving along a country lane is, however, not a particularly valid one. Motorists may strive to drive faster in order to make time gains but the same can hardly be said about speech — speakers do not

communicate to save time and break speed records, there is no distance to be covered, and not necessarily any points A and B. There might, however, be plenty of digressions.

What then is the value of speech rate to the learners? As listeners, we are sensitive to other people's SR as if we had a built-in language-speed detection device, we instantly detect whether somebody speaks fast or slowly. In our native languages this "radar" is set to a particular value which presents a kind of norm. Whilst speaking too fast may render us less intelligible (Munro and Derwing 2001), speaking too slowly might provoke impatience. Our learners are particularly at risk from the latter if, for example, they are taking part in a group communication with native speakers and want to make a contribution — if this is too slow, it may adversely affect the natural rhythm of the native communication and be viewed critically. Speeding up might, however, take its toll in giving rise to more pronunciation, prosodic and other errors (see Derwing and Munro 1997).

If we accept the notion of the built-in speech rate norm mentioned above, is it possible that this also controls the speed with which we communicate in an L2? If so, then this would be an interesting case of language transfer, which would, however, be difficult to prove as is, indeed, the case with any other examples of it. Derwing et al. (2009) in their study of 48 adult immigrants suggest that the nature of any such relationship is complex and that there can be no expectations of a straightforward link. Their study, however, only examined immigrant speakers after the first two years of linguistic development who could not yet be rated as advanced.

5.2 Fluency — frequency of unfilled pauses — results

Much attention was paid during the transcription process to the recording of pauses. Hence it was possible for the purpose of working out the pause rate to extract the annotated pauses from the transcriptions. As has been pointed out above, the length of pauses is not entirely relevant for the perception of learners' fluency and consequently I did not take it into account. The number of pauses in each task was counted using a computer script, and a UP rate was calculated (see Fig. 5–11 overleaf) as a ratio of the number of UPs per hundred words (phw).

To compare the UP frequency in the three tasks performed by the non-native speakers a one-way repeated measures ANOVA was carried out. A significant effect was found for task variability, Wilk's Lambda = .52, F (2,48) = 22.07, p < .0005. A very large effect size is evidenced by the multivariate partial eta squared of .48. To determine which of the differences is statistically significant all of the tasks were compared in a paired-samples t-test (2-tailed). A statistically significant difference was found between Tasks 1 and 3, t (49) = -4.49, p < .0005. The partial eta squared of .17 indicated a large effect size. The comparison revealed a significant difference also between Tasks 2 and 3, t (49) = 5.65, p = .0005. The partial eta squared of .,25 indicates a large effect size. A comparison of Tasks 1 and 2 did not yield a significant result (t (49) = 1.2, p > .05).

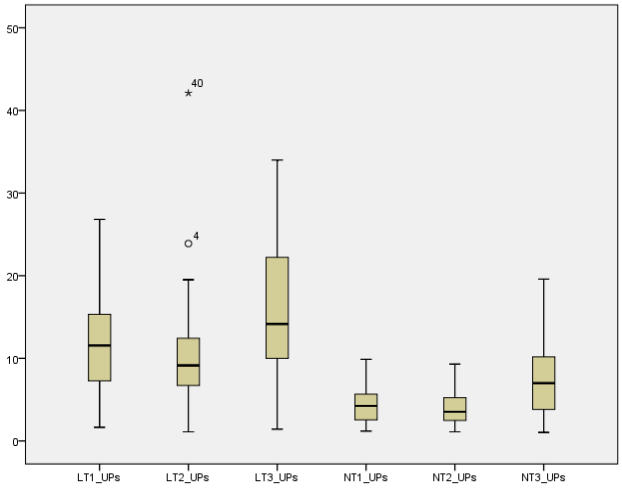

Fig. 5–11 Boxplots showing non-native (L) and native (N*) UP rates for each task separately (*T1—3). The y axis marks UP rate (number of UPs phw)*

A comparison of the three tasks performed by the native speakers using the same statistical tests proved a significant effect for task variability, Wilk's Lambda = .48, F (2,47) = 25.3 with a very large effect size (multivariate partial eta squared = .52). The paired samples t-tests revealed significant differences between all tasks. For Tasks 1 and 2, t (48) = 3.8, p < .0005; for Tasks 1 and 3, t (48) = -4.95, p < .0005; and for Tasks 2 and 3, t (48) = -6.16, p < .0005 with a large effect size expressed by the partial eta squared of .13, .2, .28 respectively.

5.2.1 Fluency — non-native-speaker frequency of unfilled pauses — results

As the variance tests showed significant difference in UP rate in the different tasks, all of the tasks are treated separately. Bar charts in Figs. 5–12—5–14 illustrate the non-native speakers' performance in the individual tasks, the bars are ordered from the lowest unfilled-pause rate to the highest.

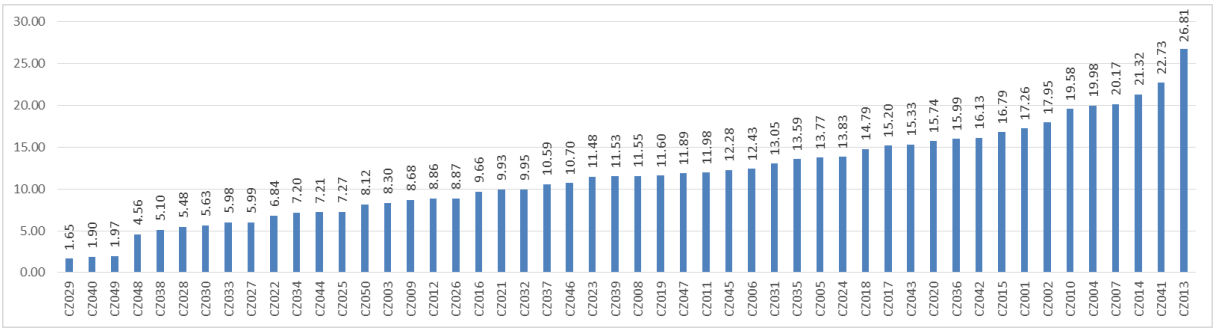

Fig. 5–12 Non-native UP rates in Task 1 in LINDSEI_CZ. The figures above the bars represent UP-rate (UPs phw)

The UP rates for Task 1 for non-native speakers range between 1.65 and 26.8 UPs phw. The mean UP rate is 11.7 (SD = 5.56) UPs phw, the median is 11.54 UPs phw.

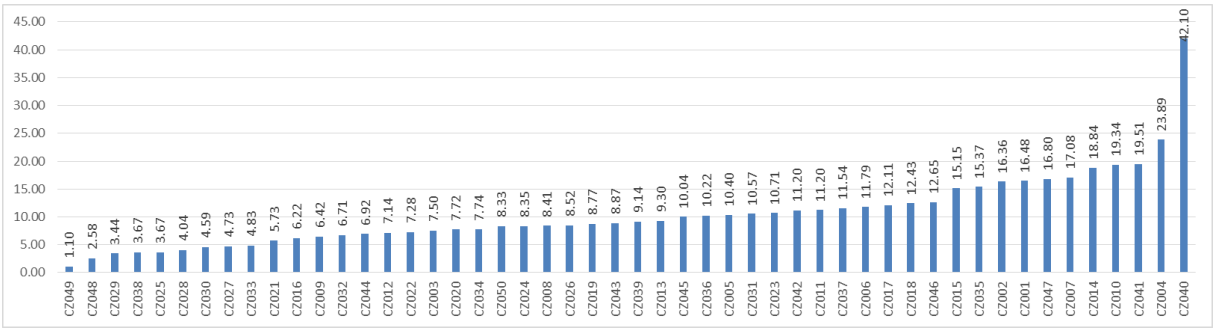

Fig. 5–13 Non-native UP rates in Task 2 in LINDSEI_CZ. The figures above the bars represent UP-rate (UPs phw)

The UP rates for Task 2 for non-native speakers range between 1.10 and 42.1 UPs phw. The mean UP rate is 10.55 (SD = 6.72) UPs phw, the median is 9 UPs phw.

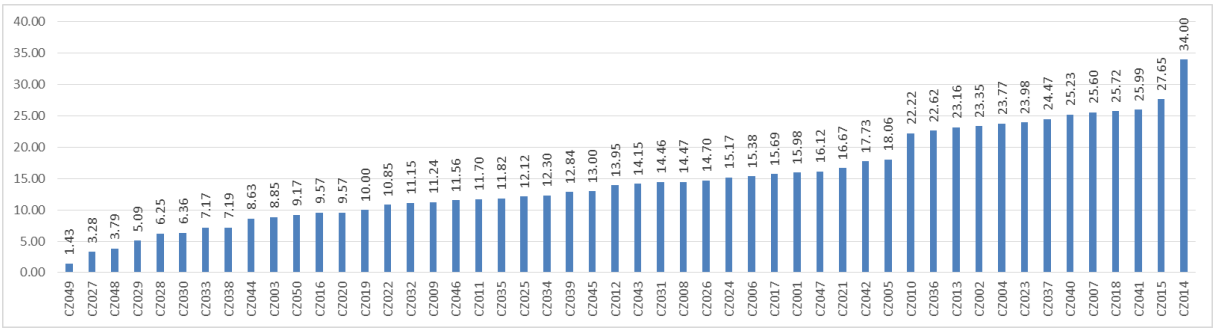

Fig. 5–14 Non-native UP rates in Task 1 in LINDSEI_CZ. The figures above the bars represent UP-rate (UPs phw)

The UP rates for Task 3 for non-native speakers range between 1.43 and 34 UPs phw. The mean UP rate is 14.9 (SD = 7.3) UPs phw, the median is 14.05 UPs phw.

5.2.2 Fluency — native-speaker frequency of UPs — results

Figs. 5–15—5–17 show bar charts illustrating the native speakers' performance in the individual tasks, the bars are ordered from the lowest unfilled-pause rate to the highest.

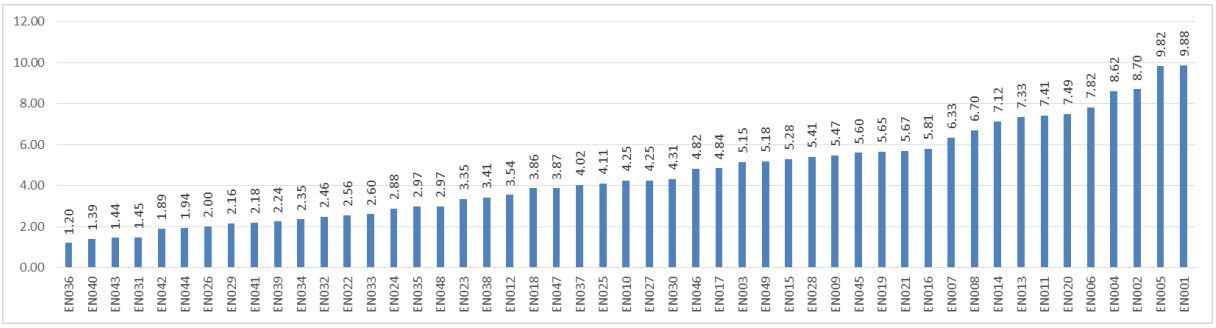

Fig. 5-15 Native UP rates in Task 1 in LOCNEC. The figures above the bars represent UP-rate (UPs phw)

The UP rates for Task 1 for native speakers range between 1.2 and 9.88 UPs phw. The mean UP rate is 4.53 (SD = 2.31) UPs phw, the median is 4.25 UPs phw.

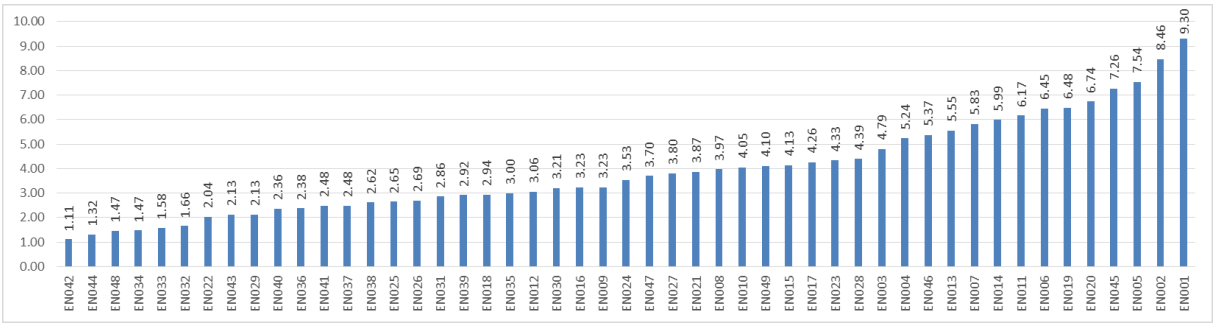

Fig. 5-16 Native UP rates in Task 2 in LOCNEC. The figures above the bars represent UP-rate (UPs phw)

The UP rates for Task 2 for native speakers range between 1.11 and 9.3 UPs phw. The mean UP rate is 3.92 (SD = 1.95) UPs phw, the median is 3.53 UPs phw.

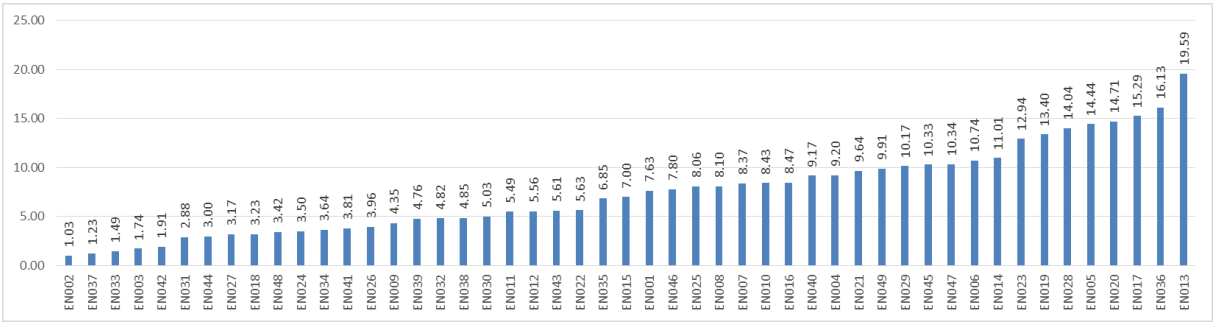

Fig. 5-17 Native UP rates in Task 3 in LOCNEC. The figures above the bars represent UP-rate (UPs phw)

The UP rates for Task 3 for native speakers range between 1.03 and 19.59 UPs phw. The mean UP rate is 7.47 (SD = 4.43) UPs phw, the median is 7 UPs phw.

5.2.3 Fluency — unfilled-pause frequency — analysis

To compare the relative frequency of UPs between LINDSEI_CZ and LOCNEC log-likelihood values were calculated as these allow comparisons between corpora of different sizes

without these being normalized first (see Table 5–2). The learners were found to overuse UPs significantly in all tasks ($p < .0001$) which confirms fluency hypothesis number 3. The frequency of UPs has a potential to distinguish native from non-native speech (cf. Riazantseva 2001).

Table 5–2 Comparison of raw frequency and log-likelihood values of UPs for LINDSEI_CZ and LOCNEC

	Task 1	**Task 2**	**Task 3**
Size in words LINDSEI	40360	43103	12525
UP count LINDSEI	4576	4341	1826
Size in words LOCNEC	43399	69269	7075
UP count LOCNEC	1808	2668	456
G^2	1450.27	1587.61	281.94
p	< .0001	< .0001	< .0001

The scatterplots and boxplots in Figs. 5–18 and 5–19 show that not only do the native speakers use fewer UPs in all tasks but they are also much more homogeneous as a group. It would appear that native speakers do not rely on UPs as a planning strategy as much as at least some of the advanced learners do. In native speech UPs are more likely to be used at natural boundaries between constituents rather than within them (cf. Götz 2013).

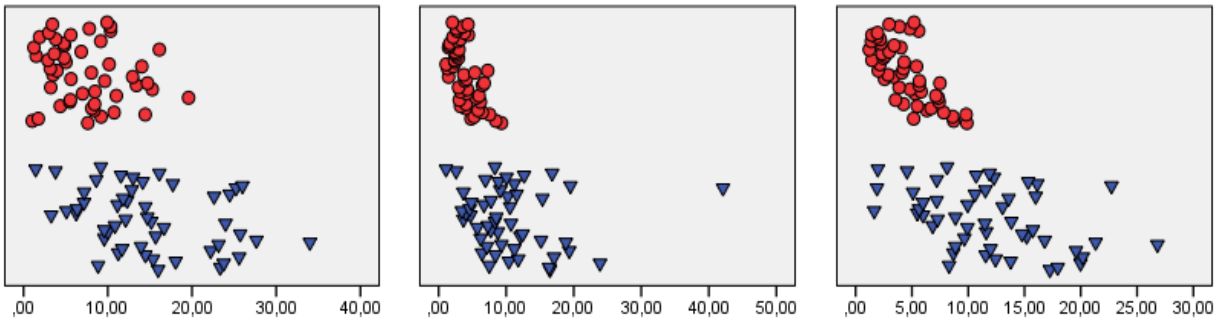

Fig. 5–18 Scatterplots showing the comparison of UP frequency for LINDSEI_CZ (the blue triangles) and LOCNEC (the red circles) in the three tasks. The x axis marks relative UP frequency (in UPs phw)

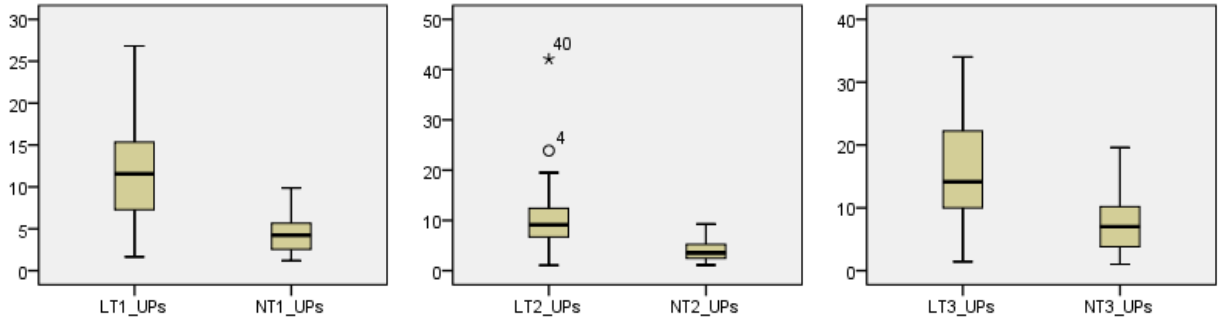

Fig. 5–19 Boxplots showing the comparison of dispersions of UP rates in LINDSEI_CZ and LOCNEC in the three tasks

The scatterplots also show that whilst there is an overlap between the two groups and at least some of the learners' ratios are comparable with those of the native speakers', a large proportion of the learners (62% in Task 1, 46% in Task 2, and 26% in Task 3) exceed the native speakers' values. As was evidenced in the SR results, Task 3 proved to be the most complex for both groups.

A closer inspection of the pausing patterns of the learners in the different tasks shows that the speakers have similar pausing habits in all of the tasks. It cannot be said with any degree of certainty whether these reflect their L1 speaking habits or whether they are more closely linked to their proficiency (cf. Riazantseva 2001).

Correlation tests, whose results were reported upon in Section 5–1, proved a moderate to large correlation between SR and a UP frequency, showing that slower speakers make a greater use of UPs. A small positive correlation ($r = .282$, $p < .05$) was found in Task 2 between the NNS's UP and FP rates suggesting that in this task those learners who used more UPs also used more FPs.

5.3 Fluency — frequency of FPs — results

Filled pauses were highlighted in the transcription using a different colour so that fast visual localisation of them would be possible on perusing the text. Each FP was then evaluated as to whether it presented an instance of a hesitation marker or backchannelling. The latter were discarded from the count, as they are simply responses to the interviewer and not part of the process of planning stretches of continuous speech. An FP rate was then calculated (see Fig. 5–20) as a ratio of FPs per hundred words (phw).

A one-way repeated measures ANOVA was carried out to compare the FP frequency in the three tasks performed by the non-native speakers. No effect was found for task variability, Wilk's Lambda = .98, $F(2,48) = .37$, $p > .05$.

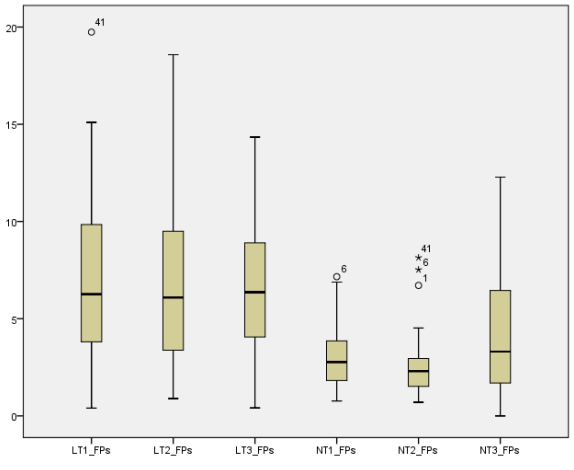

Fig. 5–20 Boxplots showing non-native (L) and native (N*) FP rates for each task separately (*T1—3). The y axis marks FP rate (number of FPs phw)*

A comparison of the three tasks performed by the native speakers using the same statistical tests proved a significant effect for task variability, Wilk's Lambda = .725, F (2,47) = 8.93 with a large effect size (multivariate partial eta squared = .275). The paired samples t-tests revealed significant differences between all tasks. For Tasks 1 and 2, t (48) = 2.75, p < .05; for Tasks 1 and 3, t (48) = -2.52, p < .05; and for Tasks 2 and 3, t (48) = -3.6, p < .05 with a moderate effect size expressed by the partial eta squared of .07, .06, .12 respectively.

5.3.1 Fluency — non-native-speaker frequency of filled pauses — results

Figs. 5–21—5–23 show bar charts illustrating the non-native speakers' performance in the individual tasks, the bars are ordered from the lowest filled-pause rate to the highest.

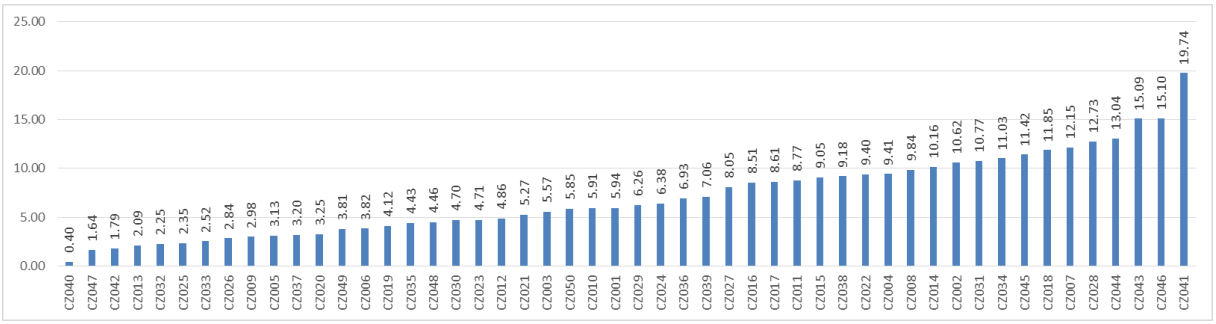

Fig. 5–21 Non-native FP rates in Task 1 in LINDSEI_CZ. The figures above the bars represent FP-rate (FPs phw)

The FP rates for Task 1 for non-native speakers range between .4 and 19.74 FPs phw. The mean FP rate is 7.06 (SD = 4.18) FPs phw, the median is 6.1 FPs phw.

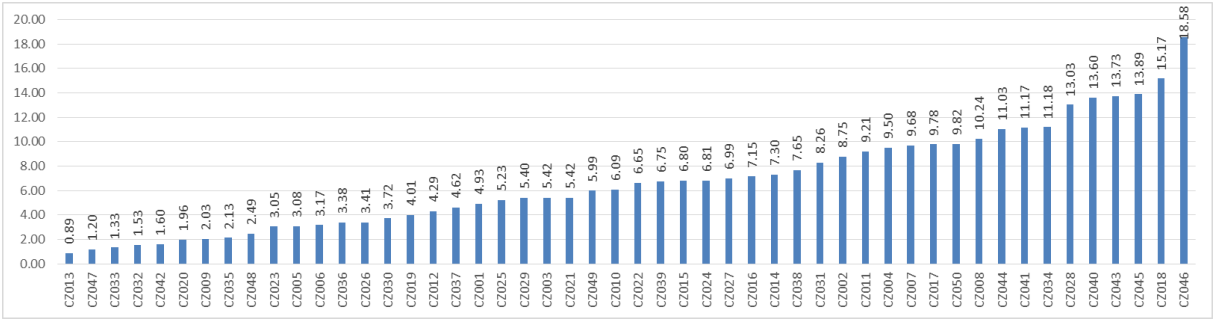

Fig. 5–22 Non-native FP rates in Task 2 in LINDSEI_CZ. The figures above the bars represent FP-rate (FPs phw)

The FP rates for Task 2 for non-native speakers range between .89 and 18.58 FPs phw. The mean FP rate is 6.78 (SD = 4.16) FPs phw, the median is 6.37 FPs phw.

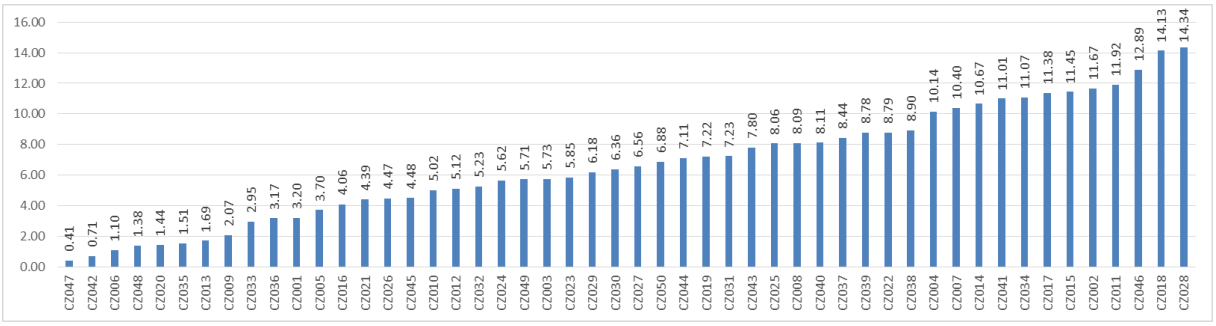

Fig. 5-23 Non-native FP rates in Task 3 in LINDSEI_CZ. The figures above the bars represent FP-rate (FPs phw)

The FP rates for Task 3 for non-native speakers range between .41 and 14.34 FPs phw. The mean FP rate is 6.69 (SD = 3.7) FPs phw, the median is 6.46 FPs phw.

5.3.2 Fluency —native-speaker frequency of filled pauses — results

Figs. 5-24—5-26 show bar charts illustrating the native speakers' performance in the individual tasks, the bars are ordered from the lowest filled-pause rate to the highest.

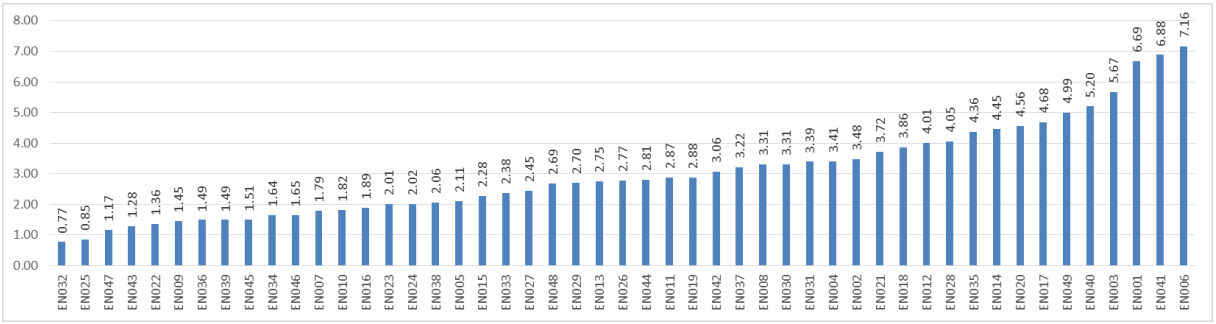

Fig. 5-24 Native FP rates in Task 1 for all LOCNEC speakers. The figures above the bars represent FP-rate (FPs phw)

The FP rates for Task 1 for native speakers range between .77 and 7.16 FPs phw. The mean FP rate is 3.03 (SD = 1.55) FPs phw, the median is 2.77 FPs phw.

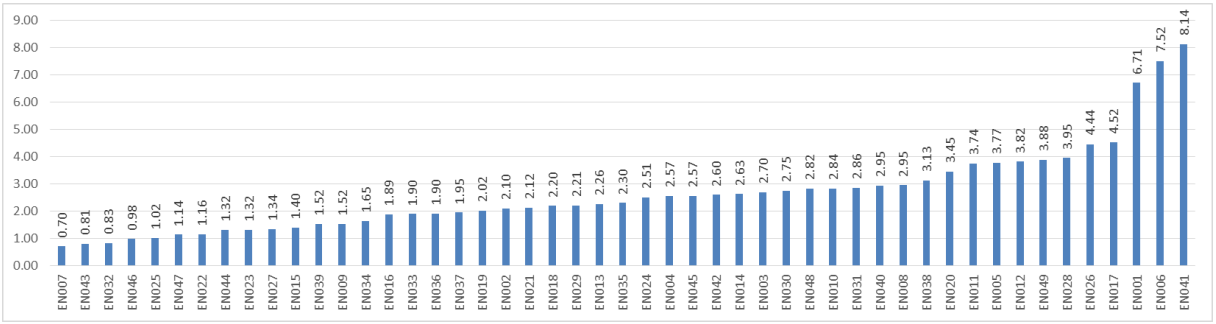

Fig. 5-25 Native FP rates in Task 2 for all LOCNEC speakers. The figures above the bars represent FP-rate (FPs phw)

The FP rates for Task 2 for native speakers range between .7 and 8.14 FPs phw. The mean FP rate is 2.64 (SD = 1.58) FPs phw, the median is 2.3 FPs phw.

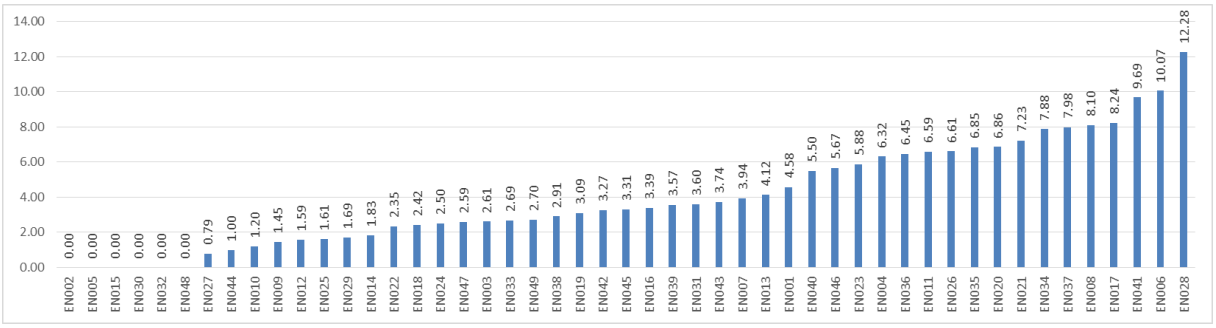

Fig. 5–26 Native FP rates in Task 3 for all LOCNEC speakers. The figures above the bars represent FP-rate (FPs phw)

The FP rates for Task 3 for native speakers range between 0 and 12.28 FPs phw. The mean FP rate is 4.02 (SD = 3.01) FPs phw, the median is 3.31 FPs phw.

5.3.3 Fluency — filled-pause frequency — analysis

The relative frequency of FPs between LINDSEI_CZ and LOCNEC was compared using log-likelihood values (see Table 5–3). The learners were found to greatly overuse FPs in all of the tasks ($p < .0001$) which confirms fluency hypothesis number 4.

Table 5–3 Comparison of raw frequency and log-likelihood values of FPs for LINDSEI_CZ and LOCNEC.

	Task 1	**Task 2**	**Task 3**
Size in words LINDSEI	40360	43103	12525
FP count LINDSEI	2717	2888	842
Size in words LOCNEC	43399	69269	7075
FP count LOCNEC	1236	1755	283
Log likelihood	681.35	1075.62	61.75
p	< .0001	< .0001	< .0001

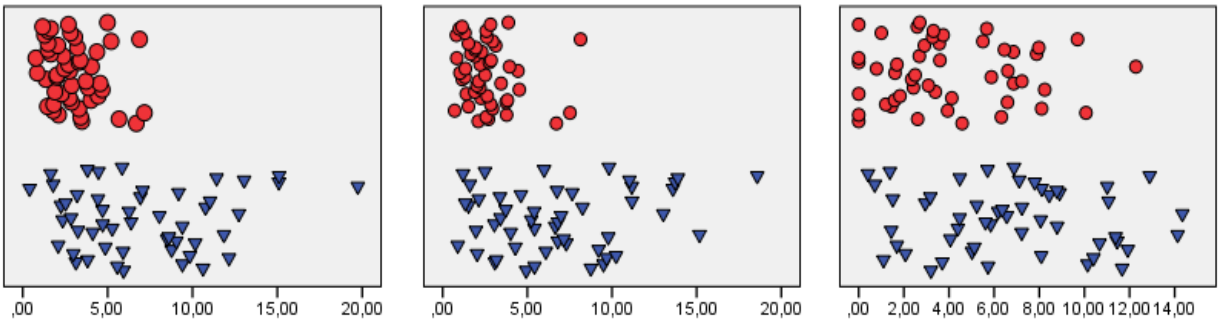

Fig. 5–27 Comparison of FP frequency for LINDSEI_CZ (the blue triangles) and LOCNEC (the red circles) in the three tasks. The x axis marks relative FP frequency (in FPs phw)

As can be seen in the scatterplots and boxplots in Figs. 5–27 (above) and 5–28 (overleaf) native speakers in Tasks 1 and 2 use fewer FPs and are also more homogeneous as a group. The learners, on the contrary produce significantly more FPs and their FP rate spans a larger range. In Task 3 we see

a somewhat different picture, the native speakers' FP rate increases and its range is comparable to that of the learners'. This would appear to imply that the design of Task 3 presents a greater challenge to spontaneous speech production even for native speakers. Whilst this is certainly evidenced by the previously measured variables (the drop in the SR and the increase in the UP rate) we can only speculate as to why there is such a disproportion in the increase between the natives and the learners.

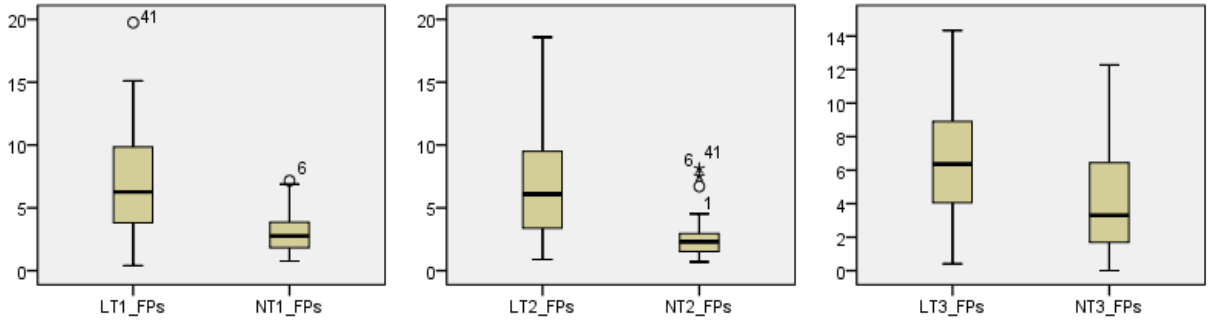

Fig. 5–28 Boxplots showing the comparison of dispersions of FP rates in LINDSEI_CZ and LOCNEC in the three tasks

As illustrated by the overlap in the scatterplots, some of the learners' performance in terms of FP rate is comparable to that of the native speakers', but a certain proportion of them (42% in Task 1, 34% in Task 2, and 6% in Task 3) exceed even the highest native-speaker FP rate.

Given the much higher frequency of FPs in non-native speech, the question arises whether most FPs produced by the learners can be functionally distinguished from UPs and whether FPs really ought to be counted as words. In my opinion, the majority of FPs used within constituents serve the same function as those UPs which are used within constituents, which is to buy time for planning. It is, however, difficult to claim with any degree of certainty what their true function is and to what extent they might actually serve a variety of functions.

Pearson correlation tests show that in LINDSEI_CZ there is a medium-strength negative correlation between speech rate and FP rate in Task 1 ($r = -.314$, $p < .05$, coefficient of determination = 10%) and in Task 3 ($r = -.309$, $p < .05$, coefficient of determination = 9.5%), but not in Task 2 ($p > .05$). A medium-strength negative correlation between SR and FP rate in Tasks 1 and 3 was found also in LOCNEC (Task 1: $r = .496$, $p < .0001$, coefficient of determination = 24.6%; Task 3: $r = .431$, $p < .05$, coefficient of determination = 18.6%), but not in Task 2 ($p > .05$). The correlation tests show that, as in the case of UP rate, slower speakers tend to produce a higher number of FPs.

The comparison of FP and UP rates in LINDSEI shows a positive small-strength correlation which is, however, statistically significant only in Task 2 ($r = .282$, $p < .05$, coefficient of determination = 8%). In LOCNEC, small to medium-strength positive correlation between UP and FP rates was found in all Tasks (Task 1: $r = .297$, $p < .05$, coefficient of determination = 8.8%; Task 2: $r = .312$, $p < .05$, coefficient of determination = 9.7%; Task 3: $r = .291$, $p < .05$, coefficient of determination = 8.5%). These results mean that at least some speakers who use more UPs tend to also use more FPs. Such a result adds further weight to the above-mentioned suggestion that FPs and UPs may in a large number of cases serve the same function and that the practice of including FPs in word counts is questionable.

The use of FPs and UPs in our sample is on the whole rather idiosyncratic. The speakers do not appear to behave in the same and predictable way in all of the tasks. It is safe to assume that the presence of both FPs and UPs indicates planning difficulties but we do not know why these have arisen. We cannot easily decide whether speakers falter because they have encountered difficulties planning content or form. Other factors may also play a role such as reduced concentration, a lack of rapport with the interviewer, insufficient interest in the topic etc. As the available metadata do not contain any information about the linguistic experience of the interviewed speakers these factors cannot be considered. It is, however, likely that, for example, public speaking experience and previous training thereof might have an influence even on spontaneous speech production.

5.4 Fluency — mean length of runs — results

The procedure in the pilot study showed that for an accurate identification of a run which can be included in the count it is essential to work not only with the transcription but also with the recording. Runs are terminated either by unfilled pauses or by the interviewer's interruptions. Interrupted runs have to be excluded as they are terminated, and thus shortened, before the speaker has a chance to complete them. However, what looks like an interruption in the transcription may not in reality always interrupt the actual run. In the following example the interviewer manages to insert the backchannelling marker in the learner's speech in such a way that it presents neither an

interruption nor an overlap. From the transcription it might appear that the marker terminates the preceding run but the recording proves the opposite.

<B> for three weeks . so we (eh) got quite acquainted with each other and (erm) she . still is really my friend </B>

<A> (mhm) </A>

<B> even though we don't write . (eh) with each other . too often . well (eh) and (eh) next year . after the . <foreign> jugend </foreign> forum . (er) we met there again </B>

It must be admitted, however, that the whole procedure is somewhat subjective and likely to be rather unreliable. Due to these limitations and the considerable time demands the technique poses I only carried out the measurements in a smaller sample of 25 students and 25 native speakers.

The pilot study showed that task design has an effect on MLRs and that differences between tasks are to be expected. Fig. 5–29 shows how the different tasks compare for the two groups of speakers in the individual tasks.

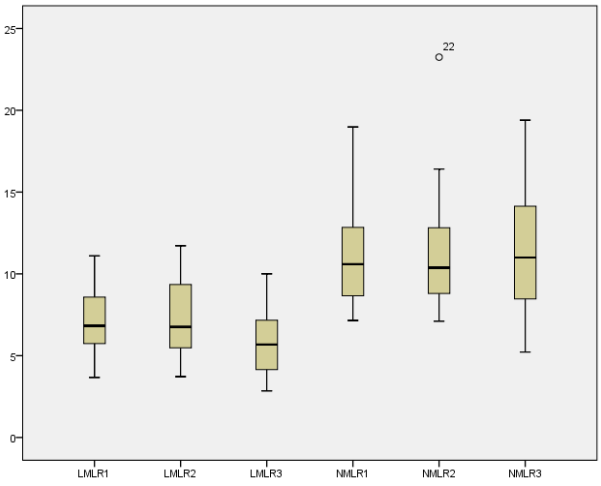

Fig. 5–29 Boxplots showing non-native (L) and native (N*) MLRs for each task separately (*T1—3). The y axis marks MLR (in words)*

To compare the MLR in the three tasks performed by the non-native speakers a one-way repeated measures ANOVA was carried out. A significant effect was found for task variability, Wilk's Lambda = .446, F (2,23) = 14.27, $p < .0005$. A large effect size is evidenced by the multivariate partial eta squared of .554. To determine which of the differences is statistically significant all of the tasks were compared in a paired-samples t-test (2-tailed). A statistically significant difference was found between Tasks 1 and 3, t (24) = 3.78, $p < .05$, and for Tasks 2 and 3, t (24) = 5.46, $p < .0005$ with a large effect size for Tasks 1 and 3 (partial eta squared of .23), and a large effect size for Tasks 2 and 3 (partial eta squared of .38). The difference between Tasks 1 and 2 did not prove to be statistically significant.

A comparison of the three tasks performed by the native speakers using the same statistical tests did not prove a significant effect for task variability ($p > .05$).

5.4.1 Fluency — non-native mean lengths of runs — results

The findings for the MLR measurements in LINDSEI_CZ are shown in Figs. 5–30—5–32.

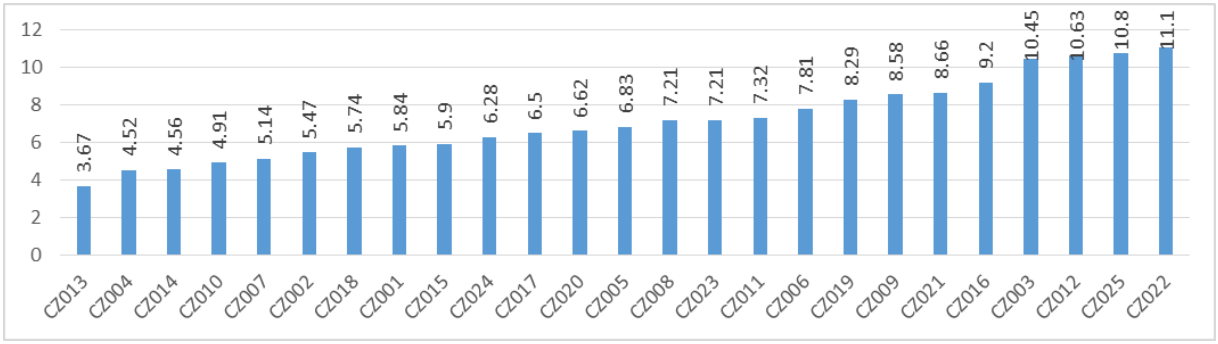

Fig. 5–30 Non-native MLRs in Task 1 for 25 LINDSEI_CZ speakers. The figures above the bars represent MLR

The MLRs for Task 1 for non-native speakers range between 3.67 and 11.1 words. The mean MLR is 7.17 (SD = 2.1), the median is 6.83 words.

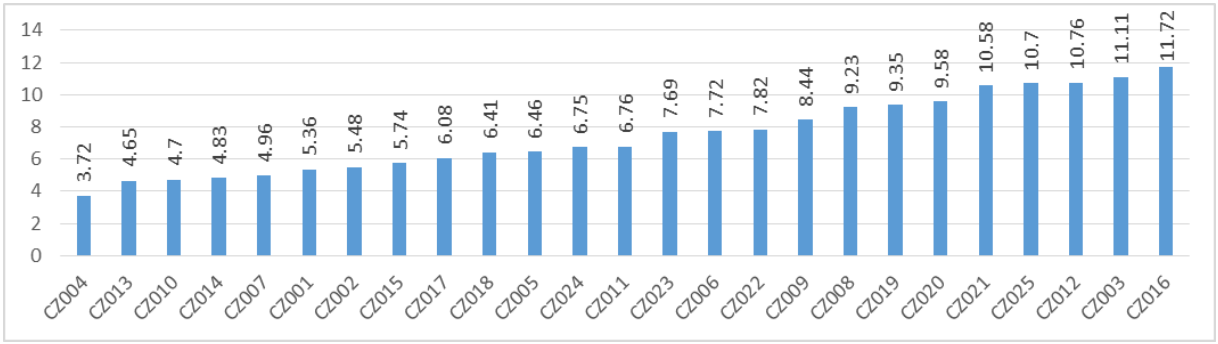

Fig. 5–31 Non-native MLRs in Task 2 for 25 LINDSEI_CZ speakers. The figures above the bars represent MLR

The MLRs for Task 2 for non-native speakers range between 3.72 and 11.72 words. The mean MLR is 7.46 (SD = 2.36), the median is 6.76 words.

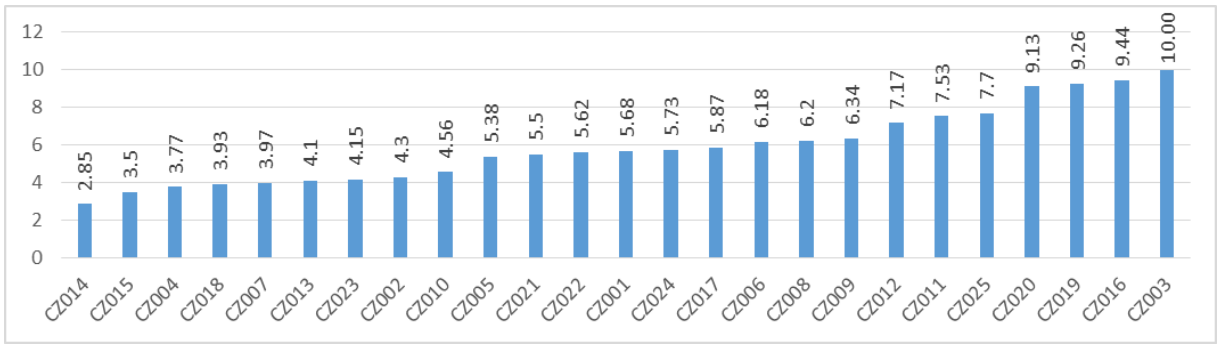

Fig. 5–32 Non-native MLRs in Task 3 for 25 LINDSEI_CZ speakers. The figures above the bars represent MLR

The MLRs for Task 3 for non-native speakers range between 2.85 and 10 words. The mean MLR is 5.91 (SD = 2.01), the median is 5.68 words.

5.4.2 Fluency —native mean lengths of runs — results

The findings for the MLR measurements in LOCNEC are shown in Figs. 5–33—5–35.

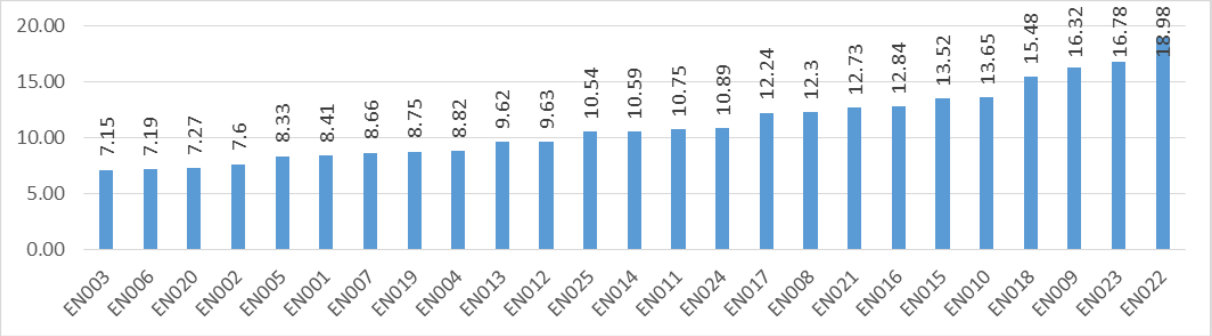

Fig. 5–33 Native MLRs in Task 1 for 25 LOCNEC speakers. The figures above the bars represent MLR (in words)

The MLRs for Task 1 for native speakers range between 7.15 and 18.98 words. The mean MLR is 11.16 (SD = 3.27), the median is 10.59 words.

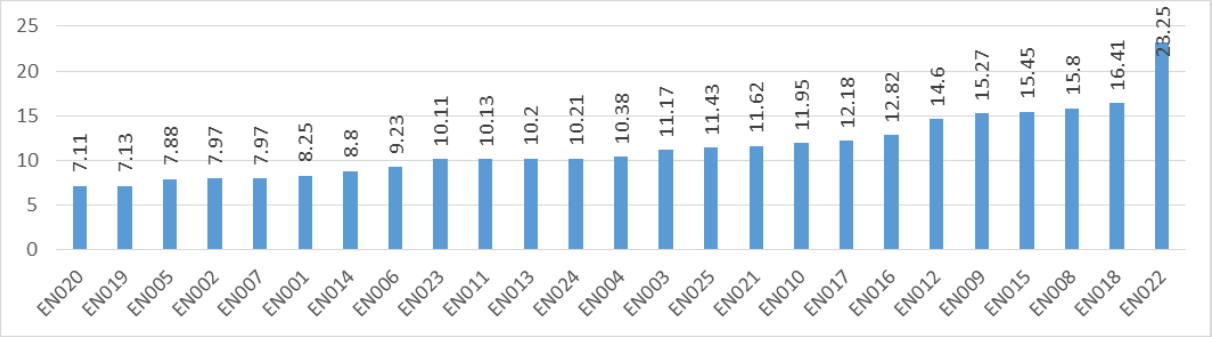

Fig. 5–34 Native MLRs in Task 2 for 25 LOCNEC speakers. The figures above the bars represent MLR (in words)

The MLRs for Task 2 for native speakers range between 7.11 and 23.25 words. The mean MLR is 11.49 (SD = 3.71), the median is 10.38 words.

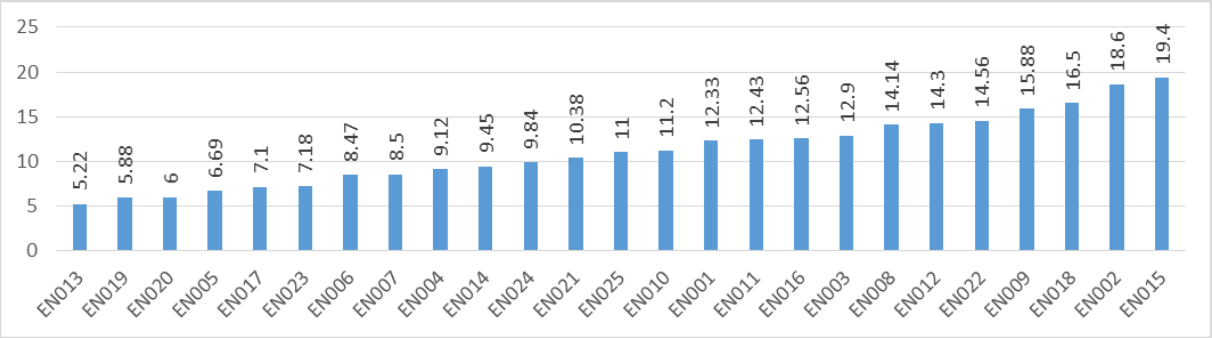

Fig. 5–35 Native MLRs in Task 3 for 25 LOCNEC speakers. The figures above the bars represent MLR (in words)

The MLRs for Task 3 for native speakers range between 5.22 and 19.4 words. The mean MLR is 11.19 (SD = 3.97), the median is 11 words.

5.4.3 Fluency — mean lengths of runs — analysis

The comparison of native and non-native MLRs reveals that native speakers produce longer runs in all tasks. The mean MLR for Task 1 in LINDSEI_CZ is 7.17 and for LOCNEC 11.16 words. For Task 2 it is 7.46 and 11.49 words respectively, and for Task 3 it is 5.91 and 11.19 words respectively. The results of t-tests prove that the differences are statistically significant ($p < .0005$). For Task 1 they are $t(24) = -5.69$, for Task 2 $t(24) = -5.04$, and for Task 3 $t(24) = -5.66$ with a large effect size (partial eta squared of .4, .35, and .4). We can therefore confirm hypothesis 5 formulated above that the learners produce shorter runs in all tasks than the native speakers. These results coincide with Götz's (2013) comparison of LOCNEC and the German subcorpus of LINDSEI.

The scatterplots and boxplots in Figs. 5–36 and 5–37 show that as regards MLR, the native speakers are much less homogeneous as a group.

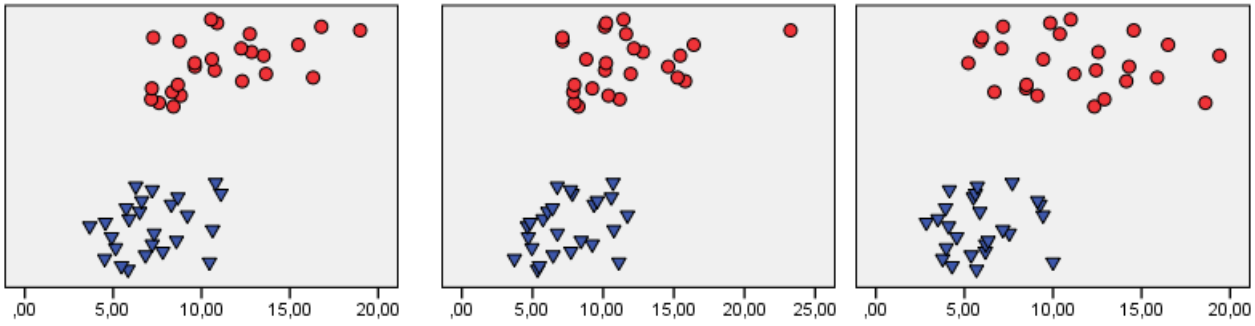

Fig. 5–36 Scatterplots showing the overlap between non-native (the blue triangles) and native (the red circles) MLRs in the three tasks. The x axis represents the MLR (in words)

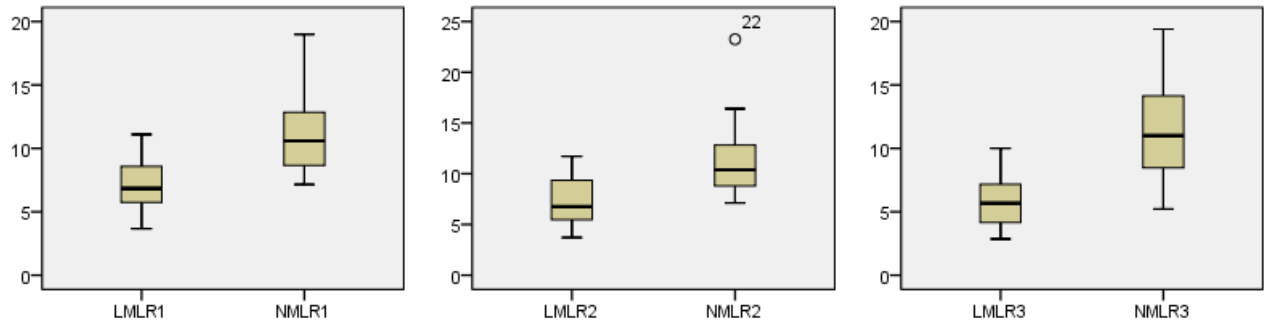

Fig. 5–37 Boxplots showing the comparisons of the dispersions of MLRs for Tasks 1—3 in LINDSEI_CZ and LOCNEC

Given the laboriousness of calculating the MLR I wanted to test its real usefulness as a separate measure. Especially as it seemed that the more UPs a speaker uses the shorter runs he

produces. In order to test this hypothesis I carried out Pearson's correlation tests. The results are shown in Table 5–4.

Table 5–4 Correlations in tasks between UP rate and MLR

	Pearson r	Strength of correlation	p	Coefficient of determination
LT1	-.936	Large	< .0005	88%
LT2	-.838	Large	< .0005	70%
LT3	-.899	Large	< .0005	81%

	Pearson r	Strength of correlation	p	Coefficient of determination
NT1	-.646	Large	< .0005	42%
NT2	-.764	Large	< .0005	58%
NT3	-.814	Large	< .0005	66%

Table 5–4 shows that there is a large negative correlation between UP rate and MLR, which means that the more UPs a speaker produces the shorter his runs are. The results are statistically significant ($p < .0005$) in all tasks for both LINDSEI_CZ (where UP rate explains between 70% and 88% of the variance in MLR) and LOCNEC (where UP rate explains between 42% and 66% of the variance in MLR). Such high correlations would appear to suggest that measuring both UP rate and MLR results in duplicity and does not offer deeper insight into fluency (cf. Norris and Ortega's (2009) warning about not grouping too many notions under one term).

5.5 Fluency — conclusion

The measurement of speech rate, unfilled pause rate, filled pause rate and mean lengths of runs for 50 non-native speakers and 49 native speakers provided a wealth of data for quantitative analyses and for comparisons of native and non-native spontaneous speech production.

Both groups of speakers proved to be fairly heterogeneous in all of the measured variables and similar to the results of fluency measurements carried out by Götz (2013) in a comparable study of the German subcorpus of LINDSEI. Both the non-native and native speakers' performance was found to be significantly affected by task design. Especially Task 3 appeared to have a large effect on the decrease of speech rate in both of the groups.

The learners proved to be slower speakers in all of the tasks, with only four speakers producing a speech rate which exceeded the mean speech rate for the native speakers. Whilst there was an overlap between the slowest native speakers and fastest learners, more than 50% of the native speakers produced more words per minute than the fastest learner.

The learners were found to significantly overuse unfilled pauses and even though there were overlaps between the two groups, the learners here were much more heterogeneous as a group with some of the speakers producing more than twice as many unfilled pauses phw. As regards the rate of use of filled pauses, the learners overused them in all of the tasks, even though in Task 3 the native speakers were found to increase their use. Except in Task 3, the native speakers were much more homogeneous as a group. The results of the measurements of the rates of use of UPs and FPs show that these variables are strong predictors of non-nativeness.

As regards the length of runs, the native speakers produce longer runs but the question arises whether this is not a predictable outcome given the large correlation between the numbers of UPs and the MLR. In this light MLR appears to be a redundant measurement (cf. Ahmadian 2012).

Correlation tests showed that slower learners have a tendency to use more unfilled and filled pauses (for FPs except in Task 2) and produce shorter speech runs.

The results of these analyses prove the hypotheses formulated in section 3.2.2.6:

1. Task design affects both native- and non-native-speaker fluency.
2. In all three tasks the learners produce fewer words per minute than the native speakers.
3. In all three tasks the learners overuse UPs.
4. In all three tasks the learners overuse FPs.
5. In all three tasks the learners produce shorter speech runs than the native speakers.

Correlation tests between accuracy and speed fluency did not prove any statistically significant correlations ($p > .05$). The following hypotheses formulated in section 3.2.2.6 may thus be rejected:

6. There is a correlation between accuracy and speed fluency.
7. There is a correlation between accuracy and breakdown fluency.

6. Conclusion

The aim of my thesis was to identify and explore the salient features of spoken English produced by Czech advanced students of English philology from the perspectives of accuracy and fluency analyses, and see how these interact, and how these learners differ from native speakers of the same age and educational background. In this last chapter I will consider the significance of the main findings and alert the reader to some if the work's weaknesses and limitations. I will then outline possible pedagogical and research implications and suggest what I see as being the most pressing directions of future research.

6.1. Findings and significance

The present study is the first large-scale description of spoken advanced-learner English produced by Czech learners. It is based on analyses of a spoken learner corpus containing almost 125,000 words in nearly 13 hours of recorded interviews. The analyses of this large amount of data revealed a surprisingly large dispersion in the measured variables, which seems to imply that the underlying proficiency of these speakers varies considerably. Subsequent division of the speakers into accuracy groups helped identify the most problematic areas as those that occur throughout in all of the accuracy bands. The most problematic aspects proved to be the use of articles and verb tenses, and especially the use of the present perfect.

As regards the learners' fluency this was compared within the group itself, showing again a large dispersion in the data, and with a parallel corpus of native speakers. The majority of the learners were found to produce speech at a considerably slower pace and with more filled and unfilled pauses and shorter speech runs. Such results possibly indicate that the underlying cognitive processes which are believed to be responsible for speech production are not yet automatized to an extent comparable with native-speaker speech production.

All of the analyses were carried out in three separate tasks. This procedure showed that task design has an effect on speech production. However, this effect was not always the same for all of the

speakers. Clearly, there are other factors besides task design which affect the dimensions of language performance.

The relationship between accuracy and fluency appears to be far from straightforward. One of the fastest speakers who also had the most native-like accent was found to have the second highest frequency of errors. And a speaker who had spent 15 years living in an English-speaking country and whose accuracy was almost absolute was found to have only an average speech rate. In fact, there did not appear to be any correlations between error rate and speed and breakdown fluency but many of the more accurate speakers also had higher speech rates. Whilst the thesis has not managed to prove such a hypothesis, these results might imply that automatic control over rule-based resources frees up attentional resources for more fluent production.

The results of the measurements are directly comparable to similar work carried out by the German LINDSEI team who, amongst their students of English philology, obtain very similar results for accuracy and fluency measures. The differences between the two nations appear in the frequency of particular error types, which indicates the possible effect of L1 transfer.

The thesis might easily create the impression that it concentrates on the shortcomings of these students rather than on their achievement which is considerable and admirable. But the present study is one of advancedness, and advancedness is a developmental stage. It is linked to progress either in time or in the acquisition of features which distinguish it from the earlier stages. The logical continuation of advanced proficiency is native-like proficiency and this does not mean to say that native-likeness is to be equated with the norm — it is simply a further step on the language acquisition continuum where no other possible goals and comparisons (e.g. other more advanced-learner stages) are readily available.

The thesis makes a contribution to the study of advancedness and deepens our understanding of it. It shows which techniques can be used to identify the common patterns and weaknesses so that pedagogical implications can be drawn (see section 6.3). It does not provide a theoretical speculation of what the speakers *can* do, as in the CEFR, but of what they really *do* do when speaking. The parameters of their performance can be used to define advanced-learner performance in English. Götz

(2013) mentions the establishment of common fluency core which would arise from comparisons of advanced learners with different L1s. The same could be done for accuracy. Such a description of advanced language has the potential to provide a gauge for judging proficiency and it may possibly also constitute a natural goal for those speakers who do not want to aim for native-like proficiency.

6.2 Limitations

Analysing spoken language easily fills the researcher with the feeling that his research contains more caveats than fool proof findings and implications. In studying spoken learner language the researcher has to confront a lack of comparable studies; the written language bias in linguistics (Linell 2005); a lack of definitions and descriptive techniques designed specifically for spoken language, and many problems arising from the very nature of spoken language which regularly makes decision-making problematic and the resulting decisions questionable. These difficulties are increased by the considerable laboriousness and time demands of spoken-language analyses. In the following paragraphs I will describe some of the limitations of my own data, the deployed techniques and my research results. I do so for two reasons. Firstly, I believe the reader ought to be aware of the existing weaknesses in the conclusions made in this thesis. Secondly, I would like to alert the reader to the fact that these problems are inherent in much learner-corpus research and present significant challenges for future research.

6.2.1 Limitations of the data

Perhaps the greatest problem of this study but also many other contemporary advanced-learner studies is in the definition of advancedness. Ortega and Byrnes (2008, 7) mention three ways of defining advancedness:[34] institutional status, information from standardised tests, the presence of

[34]Cf. Thomas (2006) who identifies four major categories of proficiency-assessment techniques – (1) Impressionistic judgment; (2) Institutional status; (3) In-house assessment; (4) Standardized test. She provides a summary of 211 studies between 2000 and 2004 and she counts 33,2% of studies that use institutional status, 23.2% use standardized tests, 19.4% in-house assessment, and 19% impressionistic judgement.

late-acquired features, and the use of sophisticated language in context. They mention that relying on one's institutional status is a widespread practice in research literature. It is simple to see why — the onus of the establishing of advancedness has been transferred to someone else. Measured advancedness would be an ideal criterion for inclusion in the corpus of advanced learners but it is logistically complex. As for late-acquired features, there does not appear to be any research which would streamline the production of a quick entry test as a criterion for inclusion. It is more than likely that in advanced learner language certain features are present that can be identified as advanced but this does not work the other way round — if we were to test only the ability to produce such features we would not have a guarantee of overall advancedness. Designing a valid test for the use of sophisticated language in context would be equally difficult.

A plausible method of ensuring the homogeneity of a spoken corpus would appear to be an introductory interview during which the interviewers would evaluate the potential participant's oral skills. The realibility of such a test would, however, be difficult to maintain.

The best way might be a post hoc evaluation of the full interview with the purpose of discarding or including. It might initially appear to be time-consuming to carry out the whole of the interview but it would obviate the need for transcription of an unsuitable text. The reliability of such a procedure would have to be tested.

As a result of this lack of clear definition of advancedness, learner corpora display a large span of proficiency. And so does LINDSEI_CZ as the results of all of the accuracy and fluency measurements showed. Whilst this provided a unique opportunity for making comparisons it possibly makes other comparisons across studies less reliable.

The second, but comparatively much less serious a problem, is the lack of data about the learners. Although they filled in a learner profile form (see Appendix 2 which contains a breakdown of all of the available metadata), the questions asked could have been more detailed. As it is we have no information, for example, about the language exams the speakers had taken, about the fact whether they use English regularly outside their university courses (i.e. at work, in a relationship with a native speaker), neither do we know anything about the way they studied English, what

methods they used, at what school and how much personal investment they made. Many more such areas could have been easily explored.

The third problem is that the intake of LINDSEI_CZ participants could have included a greater number of male speakers. The current ratio is 43 female and only 7 male speakers.

The fourth problem is somewhat more serious. The instructions provided by the organizers of the LINDSEI project are not sufficiently detailed. A lot more attention ought to have been paid when designing the corpus to the possible recording scenarios so that the final product, i.e. all of the national subcorpora, were much more comparable. Concretely, the purpose of the individual tasks ought to have been described in detail so that appropriate instructions could have been given to the learners. As it is, even within one subcorpus the nature of the tasks may vary — Task 1 is occasionally more dialogical than ought to have been the case and also the picture description in Task 3 varies in format in the various subcorpora, including in the parallel native-speaker corpus (LOCNEC). All of this makes comparisons based on these data less reliable.

The fifth problem is the transcription system. On the one hand it is very flexible and readable, on the other it lacks alignment and uses a very approximate system for transcribing overlaps. Its greatest weakness, however, is the inaccurate and highly subjective transcription of pauses.

Lastly, many of the problems connected to the complexities involved in measuring speech rates and accurately recording overlaps could have been avoided if the recordings had been made using two microphones and recorded into two tracks.

6.2.2 Limitations in the accuracy measurements

The identification of errors in speech is riddled with problems and it is thus the source of the most serious weakness in the section dealing with accuracy. Whilst many of the instances listed here were consulted with a native speaker, much more reliable results could have been obtained if more raters, and especially more native-speaker raters, had been involved in the whole of the process and

not just through private consultations of individual cases. Only in such a situation could inter-rater reliability have been calculated to provide information about the reliability of the analyses.

The Louvain error-tagging system has many advantages (especially its flexibility and its simplicity) but it fails to provide tools for a deeper classification of some of the errors, especially lexical ones. A more detailed system would be more cumbersome but it would have afforded a deeper insight into the data. It is, however, important to note that the adoption of the Louvain system rather than the development of a new one yielded results which are comparable with the other subcorpora.

Even though LINDSEI is a family of spoken corpora, the tools used for their analyses and descriptions have their roots in written language. Thus, for example, pronunciation errors are excluded although their description would significantly contribute to a deeper understanding of our advanced learners.

As regards the calculation of error rate, this is expressed as a relative frequency of a number of errors per hundred words. The definition of a word in spoken language is, however, far from straightforward. To make the study comparable with other LINDSEI subcorpora I adopted the LINDSEI default definition (which coincides with that of a graphic word) but I am aware of the fact that this is less accurate. Ideally, I would have worked with AS-units but as I explained in section 1.3 they proved extremely time-consuming to delimit in the large volume of data LINDSEI_CZ contains.

Last but not least, the main shortcoming of any error-related analysis is in its concentration on erroneous language, which somewhat detracts from the often very high quality of the language advanced learners produce.

6.2.3 Limitations in the fluency measurements

The main problems in the fluency section of the thesis stem from the problematic operationalisation of fluency. Whilst empirical research recognizes a large number of measurable variables which contribute to the impression of fluency there is not much agreement as to which the key ones are and whether in measuring these variables it is really fluency that we are measuring. Ahmadian (2012) warns that some of the measures may actually "tap the same facet of a construct"

and thus result in redundancy in measurement (Norris and Ortega 2009). Given the large volume of data LINDSEI_CZ provides, only a selection from the large number of measurable variables could be used. Thus the picture provided in my thesis only presents a picture of speed and breakdown fluency but I am fully aware of the fact that such a picture is far from complete. In my future research I hope to supplement these measures by perceptive evaluations carried out by native-speaker raters. The absence of these is yet another weakness of the present study.

On a smaller scale, my fluency measurements could have been made more accurate if a more sophisticated technique for measuring the lengths of the unfilled pauses had been available. Unfortunately, several factors precluded this, most importantly the fact the corpus is unaligned and thus obtaining any temporal measures is extremely laborious. This shortcoming may have affected the measurements of UP rates and the MLR measurements.

Another point which is somewhat problematic is the classification of filled pauses as words. Whilst this is a common practice in learner-corpus research (see e.g. Kjellmer 2003; Götz 2013) and the results of my present study are thus comparable with other studies in the field, I would be more inclined to classify filled pauses as pauses and not as words. This would than have had a direct effect on the MLR measurements both in the non-native and the native data, as filled pauses would constitute termination points for runs. This is yet another area which I hope to research in the subsequent exploration of LINDSEI.

Despite the many limitations described in the preceding paragraphs, the results of my study offer a valuable starting point for further analyses and, most importantly, they provide a wealth of data for contrastive analyses when more subcorpora have been processed using the same methodology.

6.3 Pedagogical implications[35]

In many language classrooms advanced learners are commonly left to their own devices (Cobb 2003). Their teachers believe that simply making them talk as much as possible and process vast amounts of text will do the job. Alternatively, they adopt materials for advanced language exams ignoring the fact that these exams often concentrate on developing academic skills (such is also the description of the C2 level in the CEFR) rather than aim for native-like proficiency. Fluency and accuracy development are then seen simply as by-products of exam training. When grammar and vocabulary are studied it is often the advanced and less frequent forms which receive attention and more simple features are left by the wayside.

The present study shows that advanced learners have diverse needs and shortcomings many of which seem to result from unclarities which could be easily dispelled by focussed, explicit instruction.[36] Most importantly our advanced learners still make a number of basic errors. While these might be occasionally hard to detect for the speed with which the learners produce speech, and are sometimes well hidden in the guise of sophisticated avoidance strategies, the teachers ought not to stop in their effort to identify them and take remedial action.

As general pedagogical implications for the development of fluency and accuracy are provided in sections 1.4.3 and 1.5.4, the purpose of this section is to formulate possible pedagogical implications stemming directly from the analyses carried out as part of the present thesis.

What are the weak points of our students as regards accuracy? The present study showed that this is especially the use of articles and tenses. It is not easy to identify with absolute certainty all of the reasons contributing to this situation beyond stating simply that these are notoriously difficult

[35]It has become something of a habit to conclude learner corpus studies by mentioning pedagogical implications. It is important to bear in mind that these are often lifted from other studies and whilst appearing logical and commonsensical their effectivity has rarely been empirically tested. It might therefore be more appropriate to formulate research implications and suggestions so that learner corpus research findings might eventually be verified in practice.

[36]See Leow (2015) for one of the most recent discussions of the role of explicit learning in the L2 classroom.

concepts for learners of English. They are so difficult that often even non-native teachers cannot use them accurately and most probably avoid teaching them beyond what is required in the textbooks they use, failing to provide adequate explanations. In contemporary language textbooks for advanced learners, however, articles are no longer practised and tenses are usually not dealt with systematically. These two most problematic areas thus do not receive the attention they require and the learners can make progress only through implicit learning or through self-study for which reliable materials are difficult to find. Besides, learners themselves may not be fully aware of just how problematic these features really are. The implication here is clear: the use of articles and tenses must be targeted at all levels of proficiency including the most advanced. Awareness-raising, noticing, explicit instruction, regular practice and the development of metalinguistic knowledge have been shown to be effective (e.g. Nassaji and Fotos 2011). And as standard coursebooks including those for exam preparation are short of adequate practice material, teachers are advised to supplement them with exercises from specialist grammar practice books. The teachers' main task is to provide suitable materials, encourage noticing and awareness-raising and ensure that what is difficult requires not only regular attention but also appropriate feedback.

As regards the problems with lexis, these are more diverse. The analysis showed for example gaps in the mastery of confusing pairs, which involved some fairly basic expressions, and also in the use of prepositions (especially *in*, *at* and *on*). I would endorse the findings of Zughoul's (1991) study who suggests the development of lists of problematic words. These could nowadays be easily based on learner corpus analyses. Advanced learners also ought to work with dictionaries of errors and with advanced learner's dictionaries, which often target error-prone lexical items. Teachers should not only encourage the use of such resources but also actively work with them themselves in order to raise their own awareness of common traps. The occurrence of so many "basic" errors in the speech of our advanced learners would appear to imply that this is an area teachers rather tend to neglect. Besides the Longman Dictionary of Common Errors (Turton and Heaton, 1996) which, however, urgently needs updating, other sources worth recommending are the online English

Vocabulary Profile[37] (which besides selecting useful lexical items and showing examples of use also highlights important usage details) and the Dictionary of Errors[38]. Particular attention ought to be paid to usage rather than just the development of large vocabularies as even advanced learners are found to err frequently in the use of the most frequent and seemingly simple words. Perhaps the fault lies partly with the early stages of learning when teachers were too easily inclined to condone errors which in turn became ingrained and led to the students' becoming fluent in erring.

What are the weak points of our students as regards fluency? Compared to native speakers they are slower, they produce more unfilled and filled pauses and consequently also shorter speech runs. Unlike in teaching for accuracy development where it is much easier to focus attention on specific problems, teaching for fluency development raises many questions. Section 1.4 showed that one of the soundest explanations of fluency links it to automaticity and skill acquisition. These processes rely, amongst other factors, on repetition. The learners' speech is so much slower because they need to spend more attentional resources on planning, if certain chunks are repeatedly rehearsed and as a result stored in and recalled from the mind as whole (Pawley and Syder, 1983) speech planning should become more straightforward. Significant exposure to the target language should aid this process and yet, the most accurate students in LINDSEI_CZ is the student who despite spending 15 years living in an English-speaking country is also one of the slower ones. Fluency, and even just speech rate, are clearly far more complex issues and far less understood phenomena for us to be able to formulate clear implications for classroom instruction. Götz (2013) suggests the learners ought to be taught such strategies as are used by native speakers when the flow of speech starts breaking down. She includes the use of fillers and small words (such as *like*, *you know* etc.) as strategies for buying time for planning speech but the efficiency of such strategies has not been empirically verified.

Despite there being so much discussion about the importance of teaching performance phenomena such as filled pauses I feel somewhat uncomfortable as a teacher to teach them as many people find them rather obtrusive. Perhaps teachers should instruct advanced learners to work with

[37]Available at: http://vocabulary.englishprofile.org

[38]Available at: http://www.learnglish.net/dictionary/

unfilled pauses better so that planning can happen in those moments. But, in my opinion, they should also work to reduce such filled pauses which are too frequent, too loud or too unnatural (in one of our recordings the speaker makes rather very distinct filled pauses which might be found irritating by the listeners or which might at least reduce the overall impression of the student's spoken performance).

If the use of formulaic language is conducive to success as many authors suggest (especially Wood 2012) it might be beneficial to start classroom instruction thereof in the early stages. In my own teaching practice, I have had positive experience with pre-teaching certain aspects of grammar[39] as mere phrases long before the grammatical aspects involved in the form were studied as part of the grammar syllabus. Yet again, evidence to prove the efficiency of such practice will not be easy to provide as longitudinal research in SLA is rare and as there are far too many variables at play. I am also somewhat sceptical about a solely formulaic-language based approach to language teaching as I have not yet seen a successful attempt at providing a well-organized syllabus centring around fixed expressions. Hopefully, it is only a sign of my own deficiency that I see the construction of such a syllabus a far-fetched goal, but at present I simply cannot imagine where one would start and what order one should impose on the chunks in order to organize them into a meaningful whole.

If the development of fluency is really achieved through repetition as many authors suggest (e.g. Gatbonton and Segalowitz 2005, Dörnyei 2009, Rossiter 2010)[40] the communicative language teaching practices which are still at the core of classroom teaching in this country will have to be rethought and more space allowed for repeated rehearsals, focus on form, and enhanced practice.

---

[39]E.g. such phrases as "I've never been to...", "If I were you I would..." etc.

[40]Many of the tasks suggested for fluency development here or in literature may have an immediate effect in the classroom but there is no research evidence as yet as to whether such activities have a long-term effect and actually result in long-term improvement of fluency. Bygate (1999) demonstrated the effect of repetition within a ten-week period – the repeated performance was more fluent and complex. So repetition looks promising, but this is a repetition of just one task, so perhaps this is the limitation of repetition, the ability to perform that one task more fluently. But what about spontaneous performance?

Many studies exploring the effect of stay abroad report significant improvements in fluency measures, which makes Götz (2013) suggest a compulsory stay abroad for students of English philology. I hope to live to see the day when this might become possible for all of our students. At present we might at least feel confident that the current practice of encouraging the students' participation in the Erasmus programme is a worthwhile enterprise.

My thesis also has direct pedagogical implications for the teaching of practical language at my department. It proves that the proficiency range of our students is such that they would benefit from being divided into two proficiency groups where the weaker students could be provided with more opportunities to catch up and work on more basic phenomena without slowing the more advanced students in the progress they are still to make. The results also show, as does the experience of teaching practical language, that all of the students would benefit from having more than just the two practical language courses in their first year. I can easily imagine that there could be a choice of courses so that students could choose a practical language course in each term of their study.

Another implication for the teaching at my department comes from the experience of using the present data and transcriptions in my own teaching of applied linguistics courses. Here it became clear that our students are often quite unfamiliar with many natural aspects of spoken language. In a brief experiment I carried out in autumn 2014 a group of my students were asked to find errors in the LINDSEI_CZ transcriptions. To my surprise they tagged many features which are perfectly natural in spoken English. Subsequent discussions proved that our students would benefit from linguistic courses dealing specifically with spoken grammar. Such courses would appear to be beneficial to all language-teacher trainees.

Last but not least, the present thesis appears to have some implications for language testing. In oral exams, one of the assessed components of the examinees' performance is fluency. Whilst examiners usually receive some training in this respect they might not be fully aware of the effect task design might have on performance. The results of my measurements showed that Task 3 (the story reconstruction task based on a picture description) caused considerable slowing down and an increase in filled and unfilled pausing both for the learners and for the natives. Despite the fact that

picture description is a rather unauthentic task which in real life we hardly ever have to perform, such tasks are frequently deployed at language exams. If teachers are not aware of how taxing such tasks are on attentional resources they might not accurately and justly asses the examinees' fluency.

As with teacher training mentioned above, language tester training should take into consideration the teaching of phenomena which are characteristic of spoken language so that examinees are correctly assessed. At present, in this country, there is a dearth of university courses dealing with spoken language description and analysis.

So far I have mentioned summative assessment, but teachers are more frequently involved in formative assessment, i.e. in the ongoing provision of feedback. This should also be informed as to the specificities of spoken language. Feedback should include not only information on accuracy but also on fluency. Here, it would appear useful to occasionally record students' oral performances and play them to the students in class for analysis and evaluation as we know that feedback plays an important role in learning (Ferris 2011; Nassaji and Fotos 2011).

6.4 Research implications and further research

Many of the possible research implications stem directly from the limitations outlined in section 6.4. Learner corpus research currently seems to attract a lot of attention. The field has become firmly established as a legitimate subbranch of corpus linguistics and as of spring 2015 has its own academic journal (International Journal of Learner Corpus Research) and association (Learner Corpus Association) with its biannual conference. Yet, it is still however a fairly young field and thus seems to suffer from a lack of coordinated effort with learner corpora springing up around the globe rather haphazardly. Thus the biggest challenge, in my opinion, the field seems to face is the formulation of strict standards for the design of learner corpora and detailed guidelines for quality research. The publication of the Cambridge Handbook of Learner Corpus Research (scheduled for late 2015) might be a step in the right direction. Another such step would appear to be the publication of corpus meta-analyses which should include not only the usual survey of studies but also deep analyses of the corpora themselves so that their weaknesses can be identified and learnt from.

The work with LINDSEI shows that one area which needs to receive more attention is the description of the design of the corpus and the formulation of instructions for its compilation. All parties involved should clearly understand what is expected of them and what the corpus aims to shed light upon so that when similar corpora are created (whether as part of the same project, as is the case with LINDSEI, or not) valid comparisons can be made.

The field needs replication studies and comparative studies of speakers with different L1s and yet many studies do not report on how exactly they were done, how their results were obtained. Thus it is often hard to replicate, hard to verify and sometimes the claims are thus even hard to believe. For the same reasons, monographs should, in their appendices present more detailed summaries of the data. Often we only receive summaries and we cannot look at the individual cases. Consequently, as readers we are reduced to consumers of what is presented to us and we are denied the freedom of making our own judgements. Whilst, for reasons of space, such practice is not easy to carry out in printed journal studies, their electronic versions could easily accommodate more detailed appendices. And the same applies to monographs.

Learner corpora need to collect and report on more metadata. This ought not to include only more data about the learners but also about the context in which the interviews happened. More such data will make the use of such promising explorational techniques as multifactorial feature analysis possible.

Much contemporary learner language research, including the present study, provides quantitative descriptions of language. As yet, there is a dearth of learner language theories and theorizing about the reasons for the attested phenomena. In my study I was not the first one to show that learner speech rate is incomparably slower than that of native speakers but, like others, I only succeeded in providing a partial explanation of the possible reasons (i.e. the overproduction of unfilled pauses). Future studies will have to concentrate on such areas more so that more reasons can be identified. One tool which might be of help is multi-modal learner corpora which would include not only samples of speech but also of writing and also samples of the learner's L1 production. Without such improvements we might only be painting incomplete pictures.

Research also has to focus on designing suitable and practicable tests for measuring proficiency of corpus participants or, failing that, at least criteria for pre- or post-interview evaluation of the participants' proficiency by the interviewers as I described in section 6.2. Possible directions can be drawn, for example from Callies and Götz (2015).

As regards future research related to LINDSEI_CZ, perceptive evaluations of the speakers' fluency and also of their overall proficiency by a panel of external raters will greatly contribute to a more holistic approach to fluency such as is offered e.g. by Götz (2013), and it will make it possible to pinpoint those phenomena that contribute to perceptive fluency as suggested by Fillmore (1979). The fluency measures of LINDSEI_CZ participants ought to be further analysed and especially from the perspective of the use of formulaic language and performance phenomena.

As for the accuracy, a panel of trained, native-speaker raters ought to carry out a new, complete error analysis. The individual results ought to be collated and inter-rater reliability calculated. Based on this work a catalogue of LINDSEI_CZ errors can be produced and used for comparisons with the other error-tagged subcorpora of LINDSEI. The fact that at present only two of the subcorpora are fully error-tagged points to another weakness of learner corpus research, namely the lack of international cooperation. Given the scale of the effort necessary to compile such a corpus it is surprising how little of the data have been used for truly comparative studies.

At present, LINDSEI offers a rather static view of learner language but one which could be extended into a pseudo-longitudinal corpus if the same design were used for the compilation of intermediate-level LINDSEI (and possibly other levels as well). However, proficiency would have to be carefully assessed so that a more rigorous entry criteria than simply institutional belonging (as is the case with LINDSEI) is used.

Other possible applications of LINDSEI is the study of articulation rate to complement the existing speech rate measurements and the comparison of L1 and L2 speech and articulation rates. This might reveal to what extent L2 production is affected by L1 habits.

Any future research will also have to tackle the uneasy job of assessing the learner's use of lexis in spontaneous speech production. For this, a more detailed categorization of lexical errors will have to be found. The monograph by Agustín Llach (2011) provides a sound basis in this respect.

6.5 Closing statement

Learning a foreign language requires much time and effort. In brief, it is hard work. It is the teacher's job to create such a learning environment in which the students are provided with opportunities to participate in all that contributes to the final goal while feeling the hard work to be enjoyable and worthwhile. Advanced learners often feel the need to make further progress and at the same time are deeply aware of the lack of direction on the plateau they have reached. For many of them, the enjoyment might be gained through the feeling of ongoing progress. This can only be guaranteed by teachers who are aware of the specific language needs these learners have. Learner corpus linguistics offers a vast number of opportunities to inform the work of teachers with the precise knowledge of what these needs are. It is the onus of the teacher to become familiar with the research results, as much as it is the onus of the research community to present their findings to the teacher in clear terms. Only if these two communities cooperate in this way can learner corpus research find its links to pedagogical realisations. I believe this is one of the areas in which lies the hope of future advanced learners waiting to make further progress.

Bibliography

Abutalebi, J., S. F. Cappa, and D. Perani. 2001. "The Bilingual Brain as a Function of Functional Neuroimaging." *Bilingualism: Language and Cognition* 4: 179–90.

———. 2005. "What Can Functional Neuroimaging Tell Us about the Bilingual Brain." In *Handbook of Bilingualism: Psycholinguistic Perspectives*, edited by J. Kroll and A. M. B. deGroot, 497–515. New York: Oxford University Press.

Agustín Llach, M. P. 2011. *Lexical Errors and Accuracy in Foreign Language Writing*. Second Language Acquisition 58. Bristol; Buffalo: Multilingual Matters.

Ahmadian, M. J. 2012. "The Effects of Guided Careful Online Planning on Complexity, Accuracy and Fluency in Intermediate EFL Learners' Oral Production: The Case of English Articles." *Language Teaching Research* 16 (1): 129–49. doi:10.1177/1362168811425433.

Bachman, Lyle F. 1990. *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press.

Bachman, Lyle F., and A. S. Palmer. 1996. *Language Testing in Practice: Designing and Developing Useful Language Tests*. Oxford Applied Linguistics. Oxford; New York: Oxford University Press.

Bada, E. 2006. "Pausing Preceding and Following 'That' in English." *ELT Journal* 60 (2): 125–32. doi:10.1093/elt/cci099.

Bialystok, E. 1994. "Analysis and Control in the Development of Second Language Proficiency." *Studies in Second Language Acquisition* 16 (2): 157–68.

Bialystok, E., and M. Sharwood Smith. 1985. "Interlanguage Is Not a State of Mind: An Evaluation of the Construct for Second-Language Acquisition." *Applied Linguistics* 6 (2): 101–17. doi:10.1093/applin/6.2.101.

Biber, D., S. Johansson, G. Leech, S. Conrad, and E. Finegan. 1999. *Longman Grammar of Spoken and Written English*. Harlow, England; New York: Pearson Education ESL.

Bley-Vroman, R. 1983. "The Comparative Fallacy in Interlanguage Studies: The Case of Systematicity." *Language Learning* 33: 1–17.

Bosker, H. R., A. F. Pinget, H. Quene, T. Sanders, and N. H. de Jong. 2013. "What Makes Speech Sound Fluent? The Contributions of Pauses, Speed and Repairs." *Language Testing* 30 (2): 159–75. doi:10.1177/0265532212455394.

Brand, C., and S. Götz. 2011. "Fluency versus Accuracy in Advanced Spoken Learner Language: A Multi-Method Approach." *International Journal of Corpus Linguistics* 16 (2): 255–75. doi:10.1075/ijcl.16.2.05bra.

Brumfit, C. J. 1984. *Communicative Methodology in Language Teaching: The Roles of Fluency and Accuracy*. Cambridge: Cambridge University Press.

Bulté, B., and A. Housen. 2012. "Defining and Operationalising L2 Complexity." In *Dimensions of L2 Performance and Proficiency: Complexity, Accuracy and Fluency in SLA*, edited by A. Housen, F. Kuiken, and I. Vedder. Amsterdam; Philadelphia: John Benjamins Pub. Co.

Busch, D. 1982. "Introversion-Extraversion and the EFL Proficiency of Japanese Students." *Language Learning* 32 (1): 109–32.

Bygate, M. 1999. “Task as Context for the Framing, Reframing and Unframing of Language.” *System* 27: 33–48.

———. 2001. “Effects of Task Repetition on the Structure and Control of Oral Language.” In *Researching Pedagogic Tasks, Second Language Learning, Teaching and Testing*, edited by M. Bygate, P. Skehan, and M. Swain. Harlow: Longman.

Callies, M. 2009. *Information Highlighting in Advanced Learner English: The Syntax-Pragmatics Interface in Second Language Acquisition*. Pragmatics & beyond New Series, v. 186. Amsterdam; Philadelphia: John Benjamins Pub. Co.

Callies, M., and S. Götz, eds. 2015. *Learner Corpora in Language Testing and Assessment: Prospects and Challenges*. Studies in Corpus Linguistics 70. Amsterdam; Philadelphia: John Benjamins Publishing Company.

Callies, M., and M. Paquot. 2015. “Learner Corpus Research: An Interdisciplinary Field on the Move.” *International Journal of Learner Corpus Research* 1 (1): 1–6. doi:10.1075/ijlcr.1.1.00edi.

Canale, M. 1983. “From Communicative Competence to Communicative Language Pedagogy.” In *Language and Communication*, edited by J. Richards and Schmidt, 2–28. London: Longman.

Canale, M., and M. Swain. 1980. “Theoretical Bases of Communicative Approaches to Second Language Teaching and Testing.” *Applied Linguistics* 1: 1–47.

Carlsen, C. 2012. “Proficiency Level – a Fuzzy Variable in Computer Learner Corpora.” *Applied Linguistics* 33 (2): 161–83. doi:10.1093/applin/amr047.

Carroll, J. B. 1968. “The Psychology of Language Testing.” In *Language Testing Symposium: A Psycholinguistic Approach*, edited by A. Davies, 46–69. London: Oxford University Press.

Carter, R., and M. McCarthy. 2006. *Cambridge Grammar of English: A Comprehensive Guide*. Cambridge, England; New York: Cambridge University Press.

Chambers, F. 1997. “What Do We Mean by Fluency?” *System* 25 (4): 535–44. doi:10.1016/S0346-251X(97)00046-8.

Chang, J. 2004. “Examining Models of Second Language Knowledge with Specific Reference to Relative Clauses: A Model-Comparison Approach.” Doctoral Dissertation, New York: Teachers College, Columbia University.

Chomsky, N. 1965. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT.

Christenfeld, N. 1995. “Does It Hurt to Say Um?” *Journal of Nonverbal Behavior* 19 (3): 171–86.

Clark, H., and J. E. Fox Tree. 2002. “Using Uh and Um in Spontaneous Speaking.” *Cognition* 84 (1): 73–111. doi:10.1016/S0010-0277(02)00017-3.

Coady, J., and T. N. Huckin. 1997. *Second Language Vocabulary Acquisition: A Rationale for Pedagogy*. The Cambridge Applied Linguistics Series. New York: Cambridge University Press.

Cobb, T. 2003. “Analyzing Late Interlanguage with Learner Corpora: Québec Replications of Three European Studies.” *The Canadian Modern Lan-guage Review/La Revue Canadienne Des Langues Vivantes* 59 (3): 393–423.

Collentine, J. 2004. “The Effects of Learning Contexts on Morphosyntactic and Lexical Development.” *Studies in Second Language Acquisition* 26 (2): 227–48.

Comenius, J. A. 1907. *The Great Didactic*. London: Adam and Charles Black.

Corder, S. P. 1967. "The Significance of Learners' Errors." *International Review of Applied Linguistics* 5: 161–70.

———. 1971. "Idiosyncratic Dialects and Error Analysis." *International Review of Applied Linguistics* 9: 149–59.

———. 1974. "Error Analysis." In *Techniques in Applied Linguistics*, edited by J. P. B. Allen and S. P. Corder. Language and Language Learning. London; New York: Oxford University Press.

Corder, Stephen Pit. 1981. *Error Analysis and Interlanguage*. London; New York: Oxford University Press.

Council of Europe. 2001. *Common European Framework of Reference for Languages. Learning, Teaching, Assessment*. Cambridge University Press.

Crookes, G. 1989. "Planning and Interlanguage Variation." *Studies in Second Language Acquisition* 11: 367–83.

Cucchiarini, C., H. Strik, and L. Boves. 2002. "Quantitative Assessment of Second Language Learners' Fluency: Comparisons between Read and Spontaneous Speech." *The Journal of the Acoustical Society of America* 111 (6): 2862. doi:10.1121/1.1471894.

Dagneaux, E., S. Denness, and S. Granger. 1998. "Computer-Aided Error Analysis." *System* 26 (2): 163–74. doi:10.1016/S0346-251X(98)00001-3.

Dagneaux, E., S. Denness, S. Granger, F. Meunier, J. Neff Van Aertselaer, and J. Thewissen. 2008. "Error Tagging Manual Version 1.3. Unpublished Manual." Université catholique de Louvain.

Davies, E. 1983. "Error Evaluation: The Importance of Viewpoint." *ELT Journal* 37: 304–11.

De Bot, K. 1992. "A Bilingual Production Model: Levelt's 'speaking' Model Adapted." *Applied Linguistics* 13: 1–24.

De Cock, S. 2000. "Repetitive Phrasal Chunkiness and Advanced EFL Speech and Writing." In *Corpus Linguistics and Linguistic Theory. Papers from the Twentieth International Conference on English Language Research on Computerized Corpora (ICAME 20), Freiburg Im Breisgau 1999.*, edited by C. Mair and M. Hundt, 51–68. Amsterdam: Rodopi.

———. 2012. "Spoken Learner Corpora and EFL Teaching." In *Corpus-Based Approaches to English Language Teaching*, edited by M. C. Campoy Cubillo, B. Bellés Fortuño, and M. L. Gea Valor. London; New York: Continuum.

DeKeyser, R. 1997. "Beyond Explicit Rule Learning: Automatizing Second Language Morphosyntax." *Studies in Second Language Acquisition* 19 (2): 195–221.

———. , ed. 2007. *Practice in a Second Language: Perspectives from Applied Linguistics and Cognitive Psychology*. The Cambridge Applied Linguistics Series. Cambridge; New York: Cambridge University Press.

Derwing, T. M., and M. J. Munro. 1997. "Accent, Intelligibility, and Comprehensibility." *Studies in Second Language Acquisition* 20: 1–16.

Derwing, T. M., M. J. Munro, R. I. Thomson, and M. J. Rossiter. 2009. "The Relationship between L1 Fluency and L2 Fluency Development." *Studies in Second Language Acquisition* 31 (4): 533. doi:10.1017/S0272263109990015.

Dörnyei, Z. 2009. *The Psychology of Second Language Acquisition*. Oxford; New York: Oxford University Press.

Douglas, D. 2000. *Assessing Languages for Specific Purposes*. Cambridge: Cambridge University Press.

Dulay, H. C., M. K. Burt, and S. D. Krashen. 1982. *Language Two*. New York: Oxford University Press.

Dušková, L. 1969. "On Sources of Errors in Foreign Language Learning." *International Review of Applied Linguistics* 7: 11–36.

Ellis, N. 2005. "At the Interface: Dynamic Interactions of Explicit and Implicit Language Knowledge." *Studies in Second Language Acquisition* 27 (3): 305–52.

Ellis, R. 1985. "Sources of Variability in Interlanguage." *Applied Linguistics* 6 (2): 118–31. doi:10.1093/applin/6.2.118.

———. 2003. *Task-Based Language Learning and Teaching*. Oxford: Open University Press.

———. 2008. *The Study of Second Language Acquisition*. Oxford; New York: Oxford University Press.

———. 2009. "The Differential Effects of Three Types of Task Planning on the Fluency, Complexity, and Accuracy in L2 Oral Production." *Applied Linguistics* 30 (4): 474–509. doi:10.1093/applin/amp042.

Ellis, R., and Gary Patrick Barkhuizen. 2005. *Analysing Learner Language*. Oxford Applied Linguistics. Oxford; New York: Oxford University Press.

Ellis, R., and Fangyuan Yuan. 2004. "The Effects of Planning on Fluency, Complexity, and Accuracy in Second Language Narrative Writing." *Studies in Second Language Acquisition* 26 (1): 59–84.

Ferrari, S. 2012. "A Longitudinal Study of Complexity, Accuracy and Fluency Variation in Second Language Development." In *Dimensions of L2 Performance and Proficiency: Complexity, Accuracy and Fluency in SLA*, edited by A. Housen, F. Kuiken, and I. Vedder. Amsterdam; Philadelphia: John Benjamins Pub. Co.

Ferris, D. 2011. *Treatment of Error in Second Language Student Writing*. 2nd ed. The Michigan Series on Teaching Multilingual Writers. Ann Arbor: University of Michigan Press.

Fillmore, C. J. 1979. "On Fluency." In *Individual Differences in Language Ability and Language Behavior*, edited by D. Kempler, C. J. Fillmore, and W. S. Y. Wang. New York: Academic Press.

Flynn, S., and W. O'Neil. 1988. *Linguistic Theory in Second Language Acquisition*. Dordrecht: Springer Netherlands. http://dx.doi.org/10.1007/978-94-009-2733-9.

Foster, P. 2001. "Rule and Routines: A Consideration of Their Role in Task-Based Language Production of Native and Non-Native Speakers." In *Researching Pedagogic Tasks, Second Language Learning, Teaching and Testing*, edited by M. Bygate, P. Skehan, and M. Swain. Harlow: Longman.

Foster, P., and P. Skehan. 1996. "The Influence of Planning and Task Type on Second Language Performance." *Studies in Second Language Acquisition* 18: 299–323.

Foster, P., A. Tonkyn, and G. Wigglesworth. 2000. "Measuring Spoken Language: A Unit for All Reasons." *Applied Linguistics* 21 (3): 354–75. doi:10.1093/applin/21.3.354.

Freed, B. F. 1995. “What Makes Us Think That Students Who Study Abroad Become Fluent?” In *Second Language Acquisition in a Study Abroad Context*, edited by B. F. Freed, 123–48. Philadelphia, PA: John Benjamins.

———. 2000. “Is Fluency in the Eyes (and Ears) of the Beholder?” In *Perspectives on Fluency*, edited by H. Riggenbach, 243–65. Ann Arbor: University of Michigan Press.

Fries, C. 1945. *Teaching and Learning of English as a Foreign Language*. Ann Arbor: University of Michigan Press.

Fulcher, G. 1996. “Does Thick Description Lead to Smart Tests? A Data-Based Approach to Rating Scale Construction.” *Language Testing* 13: 208–38.

Gatbonton, E., and N. Segalowitz. 2005. “Rethinking Communicative Language Teaching: A Focus on Access to Fluency.” *The Canadian Modern Lan-guage Review/La Revue Canadienne Des Langues Vivantes* 61 (3): 325–53.

Gillová, L. 2014. “Tagging a Spoken Learner Corpus.” Unpublished MA Thesis, Prague: Faculty of Arts, Charles University in Prague.

Gilquin, G., and S. De Cock. 2011. “Errors and Disfluencies in Spoken Corpora: Setting the Scene.” *International Journal of Corpus Linguistics* 16 (2): 141–72. doi:10.1075/ijcl.16.2.01gil.

Götz, S. 2007. “Performanzphänomene in Gesprochenem Lernerenglisch: Eine Korpusbasierte Pilotstudie.” *Zeitschrift Für Fremdschprachenforschung* 18 (1): 67–84.

———. 2013. *Fluency in Native and Nonnative English Speech*. Studies in Corpus Linguistics, volume 53. Amsterdam ; Philadelphia: John Benjamins Publishing Company.

———. 2015. “Tense and Aspect Errors in Spoken Learner English: Implications for Language Testing and Assessment.” In *Learner Corpora in Language Testing and Assessment*, edited by M. Callies and S. Götz. Studies in Corpus Linguistics 70. Amsterdam; Philadelphia: John Benjamins Publishing Company.

Graddol, D., J. Cheshire, and J. Swann. 1994. *Describing Language*. Buckingham: Open University Press.

Granger, S. 1994. “The Learner Corpus: A Revolution in Applied Linguistics.” *English Today* 10 (3): 25. doi:10.1017/S0266078400007665.

———. , ed. 1998. *Learner English on Computer*. Studies in Language and Linguistics. London; New York: Longman.

———. 1999. “Use of Tenses by Advanced EFL Learners: Evidence from an Error-Tagged Computer Corpus.” In *Out of Corpora: Studies in Honour of Stig Johansson.*, edited by H. Hasselgård and S. Oksefjell. Amsterdam: Rodopi.

———. 2003. “Error-Tagged Learner Corpora and CALL: A Promising Synergy.” *CALICO Journal* 20 (3): 465–80.

———. 2009. “The Contribution of Learner Corpora to Second Language Acquisition and Foreign Language Teaching: A Critical Evaluation.” In *Corpora and Language Teaching*, edited by K. Aijmer. Studies in Corpus Linguistics, v. 33. Amsterdam; Philadelphia: John Benjamins Pub. Co.

———. 2015. “Contrastive Interlanguage Analysis: A Reappraisal.” *International Journal of Learner Corpus Research* 1 (1): 7–24. doi:10.1075/ijlcr.1.1.01gra.

Granger, S., and J. Thewissen. 2005. “The Contribution of Error-Tagged Learner Corpora to the Assessment of Language Proficiency. Evidence from the International Corpus of Learner English.” In . Ottawa (Canada).

Griffiths, R. 1991. “Pausological Research in an L2 Context: A Rationale, and Review of Selected Studies.” *Applied Linguistics* 12 (4): 345–64.

Guara-Tavares, M. G. 2008. “Pre-Task Planning, Working Memory Capacity and L2 Speech Performance.” Unpublished Doctoral Thesis, Brazil: Universidade Federal de Santa Catarina.

Hamilton, R. P. 2001. “The Insignificance of Learners’ Errors: A Philosophical Investigation of the Interlanguage Hypothesis.” *Language & Communication* 21: 73–88.

Hammerly, H. 1991. *Fluency and Accuracy: Toward Balance in Language Teaching and Learning*. Multilingual Matters 73. Clevedon, Avon, England; Philadelphia: Multilingual Matters.

Han, Z. 2004. *Fossilization in Adult Second Language Acquisition*. Second Language Acquisition 5. Clevedon; Buffalo: Multilingual Matters.

Harmer, J. 2007. *The Practice of English Language Teaching*. Harlow: Pearson/Longman.

Hasselgren, A. 2002. “Learner Corpora and Language Testing: Smallwords as Markers of Learner Fluency.” In *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*, edited by S. Granger, J. Hung, and S. Petch-Tyson, 143–73. Amsterdam; Philadelphia: John Benjamins Publishing Company.

Hausner, M. 1987. “Sprechgeschwindigkeit Als Eine Funktion von Stress: Eine Fallstudie.” *Language and Style* 20: 285–311.

Hilton, H. 2008. “The Link between Vocabulary Knowledge and Spoken L2 Fluency.” *Language Learning Journal* 36 (2): 153–66.

Hincks, R. 2008. “Presenting in English or Swedish: Differences in Speaking Rate.” In *Proceedings Fonetik 2008: The XXIst Swedish Phonetics Conference*, edited by A. Eriksson and J. Lindh, 21–24. Gothenburg: Reprocentralen, Humanisten, University of Gothenburg. <http://www.ling.gu.se/konferenser/fonetik2008/papers/Proc_fonetik_2008.pdf>.

Hinkel, E. 2002. *Second Language Writers’ Text*. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Housen, A., and F. Kuiken. 2009. “Complexity, Accuracy, and Fluency in Second Language Acquisition.” *Applied Linguistics* 30 (4): 461–73. doi:10.1093/applin/amp048.

Housen, A., F. Kuiken, and I. Vedder, eds. 2012. *Dimensions of L2 Performance and Proficiency: Complexity, Accuracy and Fluency in SLA*. Amsterdam; Philadelphia: John Benjamins Publishing Company.

Hughes, A., and C. Lascaratou. 1982. “Competing Criteria for Error Gravity.” *ELT Journal* 36: 175–82.

Hulstijn, Jan Hendrik. 2015. *Language Proficiency in Native and Non-Native Speakers: Theory and Research*. Language Learning & Language Teaching, v. 41. Amsterdam; Philadelphia: John Benjamins Publishing Company.

Hunt, K. 1970. *Syntactic Maturity in School Children and Adults*. Monograph of the Society for Research into Child Development.

Hymes, D. 1972. "Competence and Performance in Linguistic Theory." In *Language Acquisition: Models and Methods*, edited by R. Huxley and Ingram, 3–28. London: Academic Press.

James, C. 1977. "Judgments of Error Gravity." *ELT Journal* 31: 116–24.

———. 1998. *Errors in Language Learning and Use: Exploring Error Analysis*. Applied Linguistics and Language Study. London; New York: Longman.

Johansson, S. 1973. "The Identification and Evaluation of Errors in Foreign Languages: A Functional Approach." In *Errata: Papers in Error Analysis*, edited by J. Svartvik. Lund: CWK Gleerup.

Kämmerer, S. 2009. "Error-Tagging Spoken Features of (learner) Language: The UCL Error Editor 'revised." In *Paper Presented at the30th Annual Conference of the International Computer Archive of Modern and Medieval English (ICAME 30)*. Lancaster.

———. 2012. "Interference in Spoken Learner Language: Scope, Dependency and Detectability." In *Proceedings from the 8th Teaching and Language Corpora Conference 2010 in Brno*, edited by J. Thomas and A. Boulton. Brno (Czech Republic): Masaryk University Press.

Khalil, K. 1985. "Communicative Error Evaluation: Native Speakers Evaluation and Interpretation of Written Errors of Arab EFL Learners." *TESOL Quarterly* 19 (2): 335–52.

Kjellmer, G. 2003. "Hesitation. In Defence of Er and Erm." *English Studies* 84 (2): 171–98.

Koponen, M., and H. Riggenbach. 2000. "Overview: Varying Perspectives on Fluency." In *Perspectives on Fluency*, edited by H. Riggenbach. Ann Arbor: University of Michigan Press.

Kormos, J. 2006. *Speech Production and Second Language Acquisition*. Cognitive Sciences and Second Language Acquisition. Mahwah, N.J: Lawrence Erlbaum Associates.

Kormos, J., and M. Dénes,. 2004. "Exploring Measures and Perceptions of Fluency in the Speech of Second Language Learners." *System* 32 (2): 145–64. doi:10.1016/j.system.2004.01.001.

Krashen, S. D. 1981. *Second Language Acquisition and Second Language Learning*. Oxford: Pergamon.

Lado, R. 1961. *Language Testing*. New York: McGraw Hill.

Lennon, P. 1984. "Retelling a Story in English." In *Second Language Productions*, edited by H. W. Dechert, D. Möhle, and M. Raupach, 50–68. Tubingen: Gunter Narr Verlag.

———. 1990a. "Investigating Fluency in EFL: A Quantitative Approach*." *Language Learning* 40 (3): 387–417. doi:10.1111/j.1467-1770.1990.tb00669.x.

———. 1990b. "The Advanced Learner at Large in the L2 Community: Developments in Spoken Performance." *International Review of Applied Linguistics* 28 (4): 309–24.

———. 1991a. "Error and the Very Advanced Learner." *International Review of Applied Linguistics in Language Teaching* 29 (1). doi:10.1515/iral.1991.29.1.31.

———. 1991b. "Error: Some Problems of Definition, Identification, and Distinction." *Applied Linguistics* 12 (2): 180–96.

———. 1996. "Getting 'Easy' Verbs Wrong at the Advanced Level," International review of applied linguistics in language teaching, 34.

———. 2000. "The Lexical Element in Spoken Second Language Fluency." In *Perspectives on Fluency*, edited by H. Riggenbach, 25–42. Ann Arbor: University of Michigan Press.

Leow, R. P. 2015. *Explicit Learning in the L2 Classroom: A Student-Centered Approach*. Second Language Acquisition Research Series. New York ; London: Routledge, Taylor & Francis Group.

Levelt, W. J. M. 1989. *Speaking: From Intention to Articulation*. Cambridge, MA: MIT Press.

———. 1999. "Language Production: A Blueprint of the Speaker." In *Neurocognition of Language*, edited by C. Brown and P. Hagoort, 83–122. Oxford: Oxford University Press.

Levis, J. M. 2008. "Pronunciation and the Assessment of Spoken Language." In *Spoken English, TESOL, and Applied Linguistics: Challenges for Theory and Practice*, edited by Rebecca Hughes. New York: Palgrave Macmillan.

Levkina, M., and R. Gilabert. 2012. "The Effects of Cognitive Task Complexity on L2 Oral Production." In *Dimensions of L2 Performance and Proficiency: Complexity, Accuracy and Fluency in SLA*, edited by A. Housen, F. Kuiken, and I. Vedder. Amsterdam; Philadelphia: John Benjamins Publishing Company.

Lewis, M. 1993. *The Lexical Approach: The State of ELT and a Way Forward*. Hove, England: Language Teaching Publications.

———. 2000. *Teaching Collocation: Further Developments in the Lexical Approach*. Hove: Language Teaching Publications.

Lewis, M., and C. Gough. 1997. *Implementing the Lexical Approach: Putting Theory into Practice*. Hove, England: Language Teaching Publications.

Lightbown, P., and N. M. Spada. 2013. *How Languages Are Learned*. Fourth edition. Oxford Handbooks for Language Teachers. Oxford: Oxford University Press.

Linell, P. 2005. *The Written Language Bias in Linguistics: Its Nature, Origins and Transformations*. London: Routledge. http://site.ebrary.com/id/10162613.

Littlewood, W. 1981. *Communicative Language Teaching: An Introduction*. New Directions in Language Teaching. Cambridge [Eng.] ; New York: Cambridge University Press.

Loban, W. 1966. "The Language of Elementary School Children. (Research Report No. 1)." *Champaign III: National Council of Teachers of English*.

Long, M. 2003. "Stabilization and Fossilization in Interlanguage Development." In *Handbook of Second Language Acquisition*, edited by C. Doughty and M. Long. Malden: Blackwell.

Mackey, A., and S. M. Gass, eds. 2012. *Research Methods in Second Language Acquisition: A Practical Guide*. 1st ed. Guides to Research Methods in Language and Linguistics. Chichester, West Sussex; Malden, Mass: Wiley-Blackwell.

McCretton, E., and N. Rider. 1993. "Error Gravity and Error Hierarchies." *International Review of Applied Linguistics in Language Teaching* 31 (3).

Meara, P. 1984. "The Study of Lexis in Interlanguage." In *Interlanguage*, edited by A. Davies, A. P. R Howatt, and C. Criper, 225–39. Edinburgh: Edinburgh University Press.

Mehnert, U. 1998. "The Effects of Different Lengths of Time for Planning on Second Language Performance." *Studies in Second Language Acquisition* 20: 52–83.

Meisel, J., H. Clahsen, and M. Pienemann. 1981. "On Determining Developmental Stages in Natural Second Language Acquisition." *Studies in Second Language Acquisition* 3: 109–35.

Miller, J. E., and R. Weinert. 2009. *Spontaneous Spoken Language: Syntax and Discourse*. Oxford; New York: Oxford University Press.

Milton, J. 2009. *Measuring Second Language Vocabulary Acquisition*. Second Language Acquisition. Bristol; Buffalo: Multiligual Matters.

Möhle, D. 1984. "A Comparison of the Second Language Speech Production of Different Native Speakers." In *Second Language Productions*, edited by H. W. Dechert, D. Möhle, and M. Raupach, 50–68. Tübingen: Narr.

Mora, J. C. 2006. "Age Effect on Oral Fluency Development." In *Age and the Rate of Foreign Language Learning*, edited by C. Muñoz. Clevedo; Buffalo: Multilingual Matters. http://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=164454.

Mukherjee, J. 2009. "The Grammar of Conversation in Advanced Spoken Learner English: Learner Corpus Data and Language-Pedagogical Implications." In *Corpora and Language Teaching*, edited by K. Aijmer. Studies in Corpus Linguistics, v. 33. Amsterdam; Philadelphia: John Benjamins Publishing Company.

Müller, S. 2005. *Discourse Markers in Native and Non-Native English Discourse*. Pragmatics & beyond, new ser. v. 138. Amsterdam; Philadelphia: John Benjamins Publishing Company.

Munro, M. J., and T. M. Derwing. 2001. "Modelling Perceptions of the Accentedness and Comprehensibility of L2 Speech: The Role of Speaking Rate." *Studies in Second Language Acquisition* 23: 451–68.

Nassaji, H., and S. Fotos. 2011. *Teaching Grammar in Second Language Classrooms: Integrating Form-Focused Instruction in Communicative Context*. New York: Routledge.

Nation, I. S. P. 1999. "Fluency and Accuracy." In *Concise Encyclopedia of Educational Linguistics*, edited by B. Spolsky and R. E. Asher, 611–12. Amsterdam; New York: Elsevier.

———. 2001. *Learning Vocabulary in Another Language*. The Cambridge Applied Linguistics Series. Cambridge; New York: Cambridge University Press.

———. 2009. *Teaching ESL/EFL Listening and Speaking*. ESL & Applied Linguistics Professional Series. New York: Routledge.

Norris, J. M., and L. Ortega. 2003. "Defining and Measuring SLA." In *Handbook of Second Language Acquisition*, edited by C. Doughty and M. Long, 716–61. Malden: Blackwell.

———. 2009. "Towards an Organic Approach to Investigating CAF in Instructed SLA: The Case of Complexity." *Applied Linguistics* 30 (4): 555–78. doi:10.1093/applin/amp044.

O'Connell, D. C., and S. Kowal. 2004. "The History of Research on the Filled Pause as Evidence of the Written Language Bias in Linguistics (Linell, 1982)." *Journal of Psycholinguistic Research* 33 (6): 459–74.

Ortega, L. 1999. "Planning and Focus on Form in L2 Oral Performance." *Studies in Second Language Acquisition* 21: 109–48.

Ortega, L., and H. Byrnes, eds. 2008. *The Longitudinal Study of Advanced L2 Capacities*. Second Language Acquisition Research. New York; London: Routledge.

Osborne, J. 2007. "Why Do They Keep Making the Same Mistakes? Evidence for Error Motivation in a Learner Corpus." In *Corpora and ICT in Language Studies: PALC 2005*, edited by J. Waliński, K. Kredens, and S. Goźdź-Roszkowski. Łódź Studies in Language, v. 13. Frankfurt am Main ; New York: Lang.

Pallotti, G. 2009. "CAF: Defining, Refining and Differentiating Constructs." *Applied Linguistics* 30 (4): 590–601. doi:10.1093/applin/amp045.

Pawley, A., and F. H. Syder. 1983. "Two Puzzles for Linguistic Theory: Nativelike Selection and Nativelike Fluency." *Language and Communication* 191. http://folders.nottingham.edu.cn/staff/webct_local_serve/Vocabulary%20-%20Teaching%20and%20Learning%20I%20(ES4159)/Vocab/pawleysyder.pdf.

Polio, C. G. 1997. "Measures of Linguistic Accuracy in Second Language Writing Research." *Language Learning* 47 (1): 101.

Purpura, J. E. 2004. *Assessing Grammar*. Cambridge: Cambridge University Press.

Ramig, L. A. 1983. "Effects of Physiological Aging on Speaking and Reading Rates." *Journal of Communication Disorders* 16 (3): 217–26.

Raupach, M. 1980. "Temporal Variables in First and Second Language Speech Production." In *Temporal Variables in Speech*, edited by H. W. Dechert and M. Raupach, 263–70. The Hague: Mouton.

———. 1983. "Analysis and Evaluation of Communication Strategies." In *Strategies in Interlanguage Communication*, edited by C. Faerch and G. Kasper, 199–209. London: Longman.

Reason, J. 1990. *Human Error*. 2009th ed. Cambridge: Cambridge University Press.

Riazantseva, A. 2001. "Second Language Proficiency and Pausing." *Studies in Second Language Acquisition* 23: 497–526.

Richards, J. 1971. "A Non-Contrastive Approach to Error Analysis." *ELT Journal* 25 (2): 204–19.

Richards, Jack C., ed. 1973. *Error Analysis. Perspectives on Second Language Acquisition.* Harlow: Longman.

Riggenbach, H. 1991. "Toward an Understanding of Fluency: A Microanalysis of Nonnative Speaker Conversations." *Discourse Processes* 14: 423–41.

———. , ed. 2000. *Perspectives on Fluency*. Ann Arbor: University of Michigan Press.

Ringbom, H. 1987. *The Role of the First Language on the Foreign Language Learning*. Clevedon: Multilingual Matters.

Robinson, P. 2001a. "Task Complexity, Cognitive Resources, and Syllabus Design: A Triadic Framework for Examining Task Influences on SLA." In *Cognition and Second Language Instruction*, edited by P. Robinson, 287–318. Cambridge: Cambridge University Press.

———. 2001b. "Task Complexity, Task Difficulty, and Task Production: Exploring Interactions in a Componential Framework." *Applied Linguistics* 22 (1): 27–57. doi:10.1093/applin/22.1.27.

Robinson, P., T. Cadierno, and Y. Shirai. 2009. "Time and Motion: Measuring the Effects of the Conceptual Demands of Tasks on Second Language Speech Production." *Applied Linguistics* 30 (4): 533–54. doi:10.1093/applin/amp046.

Rossiter, M. J., T. M. Derwing, L. G. Manimtim, and R. I. Thomson. 2010. "Oral Fluency: The Neglected Component in the Communicative Language Classroom." *Canadian Modern Language Review/ La Revue Canadienne Des Langues Vivantes* 66 (4): 583–606. doi:10.3138/cmlr.66.4.583.

Rühlemann, C. 2006. "Coming to Terms with Conversational Grammar: 'Dislocation' and 'Dysfluency.'" *International Journal of Corpus Linguistic* 11 (4): 385–409.

Rundell, M. 2007. *Macmillan English Dictionary for Advanced Learners*. Oxford: Macmillan.

Salaberry, M. R., and N. Lopez-Ortega. 1998. "Accurate L2 Production across Language Tasks: Focus on Form, Focus on Meaning, and Communicative Control." *The Modern Language Journal* 82 (4): 514–32. doi:10.1111/j.1540-4781.1998.tb05539.x.

Sangarun, J. 2005. "The Effects of Focusing on Meaning and Form in Strategic Planning." In *Planning and Task Performance in a Second Language*, edited by Rod Ellis. Amsterdam: John Benjamins Publishing Company.

Sasaki, M. 2002. "Building an Empirically-Based Model of EFL Learners' Writing Processes." In *Studies in Writing*, edited by S. Ransdell and M. Barbier, 101–22. New Directions in Research in L2 Writing, Vol. 11. Dordrecht: Kluwer Academic Publishers.

Schachter, J., and M. Celce-Murcia. 1977. "Some Reservations Concerning Error Analysis." *TESOL Quarterly* 11: 141–51.

Schmidt, R. 1992. "Psychological Mechanisms Underlying Second Language Fluency." *Studies in Second Language Acquisition* 14: 357–85.

———. 2001. "Attention." In *Cognition and Second Language Instruction*, edited by P. Robinson. New York: Cambridge University Press.

Schmitt, N. 2010. *Researching Vocabulary: A Vocabulary Research Manual*. Research and Practice in Applied Linguistics. Houndmills, Basingstoke, Hampshire ; New York, NY: Palgrave Macmillan.

Schumann, J. H. 1978. *The Pidginization Process: A Model for Second Language Acquisition*. Series in Sociolinguistics. Rowley, Mass: Newbury House Publishers.

Schwartz, B. D. 1997. "On the Basis of the Basic Variety..." *Second Language Research* 13 (4): 386–402. doi:10.1191/026765897670070857.

Scrivener, J. 2011. *Learning Teaching: The Essential Guide to English Language Teaching*. Oxford: Macmillan Education.

Segalowitz, N. 2000. "Automaticity and Attentional Skill in Fluent Performance." In *Perspectives on Fluency*, edited by H. Riggenbach. Ann Arbor: University of Michigan Press.

———. 2010. *Cognitive Bases of Second Language Fluency*. Cognitive Science and Second Language Acquisition Series. New York: Routledge.

Segalowitz, N., and B. F. Freed. 2004. "Context, Contact, and Cognition in Oral Fluency Acquisition: Learning Spanish in at Home and Study Abroad Contexts." *Studies in Second Language Acquisition* 26 (2): 173–99.

Selinker, L. 1972. "Interlanguage." *International Review of Applied Linguistics* 10: 209–31.

Skehan, P. 1996. “Second Language Acquisition and Task-Based Instruction.” In *Challenge and Change in Language Teaching*, edited by J. Willis and D. Willis, 17–30. Oxford: Heinemann.

———. 1998. *A Cognitive Approach to Language Learning*. Oxford: Oxford University Press.

———. 2001. “Tasks and Language Performance Assessment.” In *Researching Pedagogic Tasks, Second Language Learning, Teaching and Testing*, edited by M. Bygate and M. Swain. Harlow: Longman.

———. 2003. “Task-Based Instruction.” *Language Teaching* 36 (1): 1–14.

———. 2009a. “Lexical Performance by Native and Non-Native Speakers on Language Learning Tasks.” In *Vocabulary Studies in First and Second Language Acquisition: The Interface between Theory and Application*.

———. 2009b. “Modelling Second Language Performance: Integrating Complexity, Accuracy, Fluency, and Lexis.” *Applied Linguistics* 30 (4): 510–32. doi:10.1093/applin/amp047.

Skehan, P., and P. Foster. 1997. “Task Type and Task Processing Conditions as Influences on Foreign Language Performance.” *Language Teaching Research* 1: 185–211.

———. 1999. “The Influence of Task Structure and Processing Conditions on Narrative Retellings.” *Language Learning* 49 (1): 93–120.

Spillner, B. 1991. *Error Analysis: A Comprehensive Bibliography*. Amsterdam: John Benjamins Publishing Company.

Stenson, B. 1974. “Induced Errors.” In *New Frontiers in Second Language Learning*, edited by J. Schumann and B. Stenson. Rowley: Newbury House.

Sternberg, R. J., K. Sternberg, and J. S. Mio. 2009. *Cognitive Psychology*. Belmont, CA: Wadsworth/Cengage Learning.

Štindlová, B. 2011. “Evaluace Chybové Anotace v Žákovském Korpusu Češtiny [Evaluation of Error Mark-up in a Learner Corpus of Czech].” Disertační práce, Praha: Filosofická fakulta Univerzity Karlovy.

Svartvik, J. 1973. *Errata: Papers in Error Analysis*. Lund: CWK Gleerup.

Swan, M., and B. Smith, eds. 2001. *Learner English: A Teacher’s Guide to Interference and Other Problems*. 2nd ed. Cambridge Handbooks for Language Teachers. Cambridge; New York: Cambridge University Press.

Tauroza, S., and D. Allison. 1990. “Speech Rates in British English.” *Applied Linguistics* 11 (1): 90–105.

Tavakoli, P., and P. Skehan. 2005. “Strategic Planning, Task Structure and Performance Testing.” In *Planning and Task Performance in a Second Language*, edited by R. Ellis, 239–77. Amsterdam: John Benjamins.

Taylor, G. 1986. “Errors and Explanations.” *Applied Linguistics* 7 (2): 144–66.

Thewissen, J. 2013. “Capturing L2 Accuracy Developmental Patterns: Insights From an Error-Tagged EFL Learner Corpus: Capturing L2 Accuracy Developmental Patterns.” *The Modern Language Journal* 97 (S1): 77–101. doi:10.1111/j.1540-4781.2012.01422.x.

Thomas, M. 1994. “Assessment of L2 Proficiency in Second Language Acquisition Research.” *Language Learning* 44: 307–36.

———. 2006. “Research Synthesis and Historiography: The Case of Assessment of Second Language Proficiency.” In *Synthesizing Research on Language Learning and Teaching*, edited by J. M. Norris and L. Ortega, 279–98. Language Learning and Language Teaching, v. 13. Amsterdam; Philadelphia, Pa: John Benjamins Publishing Company.

Tottie, G. 2012. “Planning Your Turns: Uh, Um and Speaker Change.” In *The Cambridge Handbook of Pragmatics*, edited by K. Allan and K. Jaszczolt. Cambridge Handbooks in Language and Linguistics. Cambridge; New York: Cambridge University Press.

Tottie, G., and S. Svalduz. 2009. “Er/uh, Erm/um: Planners in Varieties and Registers of English.” In *Presented at ICAME 30*. Lancaster University.

Towell, R. 1987. “Variability and Progress in the Language Development of Advanced Learners of a Foreign Language.” In *Second Language Acquisition in Context*, edited by R. Ellis, 113–27. Toronto: Prentice-Hall.

———. 2002. “Relative Degrees of fluency: A Comparative Case Study of Advanced Learners of French.” *International Review of Applied Linguistics in Language Teaching* 40: 117–50.

———. 2012. “Complexity, Accuracy and Fluency in Second Language Acquisition Research.” In *Dimensions of L2 Performance and Proficiency: Complexity, Accuracy and Fluency in SLA*, edited by A. Housen, F. Kuiken, and I. Vedder, 47–69. Amsterdam; Philadelphia: John Benjamins Publishing Company.

Towell, R., R. Hawkins, and N. Bazergui. 1996. “The Development of Fluency in Advanced Learners of French.” *Applied Linguistics* 17 (1): 84–119.

Tracy-Ventura, N., and F. Myles. 2015. “The Importance of Task Variability in the Design of Learner Corpora for SLA Research.” *International Journal of Learner Corpus Research* 1 (1): 58–95. doi:10.1075/ijlcr.1.1.03tra.

Trenchs-Parera, M. 2009. “Effects of Formal Instruction and a Stay Abroad on the Acquisition of Native-like Oral Fluency.” *The Canadian Modern Lan-guage Review/La Revue Canadienne Des Langues Vivantes* 65 (3): 365–93.

Turton, N. D., and J. B Heaton. 1996. *Longman Dictionary of Common Errors*. Harlow: Longman.

Ur, P. 2012. *A Course in English Language Teaching*. 2nd ed. Cambridge; New York: Cambridge University Press.

Valette, R. 1991. “Proficiency and the Prevention of Fossilization – an Editorial.” *Modern Language Journal* 75 (3): 326–28.

Van Daele, S., A. Housen, M. Pierrard, and L. De Bruyn. 2006. “The Effect of Extraversion on Oral L2 Proficiency.” *EUROSLA Yearbook* 6 (1): 213–36.

VanPatten, B. 1990. “Attending to Form and Content in the Input.” *Studies in Second Language Acquisition* 12 (3): 287–301.

Wardhaugh, R. 1970. “The Contrastive Analysis Hypothesis.” *TESOL Quarterly* 4: 123–30.

Watkins, P. 2005. *Learning to Teach English*. Peaslake, England: Delta Publishing.

Weir, C. J. 1997. “The Testing of Reading.” In *Encyclopedia of Language and Education,* edited by C. Clapham and Corson, 39–49. Language Testing and Assessment, Vol. 7. Dordrecht: Kluwer Academic Publishers.

Wendel, J. 1997. "Planning and Second Language Narrative Production." Unpublished Doctoral Dissertation, Japan: Temple University.

Whiteside, S. P. 1996. "Temporal-Based Acoustic-Phonetic Patterns in Read Speech: Some Evidence for Speaker Sex Differences." *Journal of the International Phonetic Association* 26 (01): 23. doi:10.1017/S0025100300005302.

Widdowson, H. G. 1978. *Teaching Language as Communication*. Oxford: Oxford University Press.

Wigglesworth, G. 1997. "An Investigation of Planning Time and Proficiency Level on Oral Test Discourse." *Language Testing* 14 (1): 21–44.

Wolfe-Quintero, K., S. Inagaki, and H.-Y. Kim. 1998. *Second Language Development in Writing: Measures of Fluency, Accuracy, & Complexity*. Honolulu: Second Language Teaching & Curriculum Center, University of Hawaii at Manoa.

Wood, D. 2012. *Formulaic Language and Second Language Speech Fluency: Background, Evidence and Classroom Applications*. London; New York: Continuum.

Wray, A. 2000. "Formulaic Sequences in Second Language Teaching: Principle and Practice." *Applied Linguistics* 21 (4): 463–89. doi:10.1093/applin/21.4.463.

Yuan, F., and R. Ellis. 2003. "The Effects of Pre-Task and on-Line Planning on Fluency, Complexity and Accuracy in L2 Monologic Oral Production." *Applied Linguistics* 24 (1): 1–27.

Zughoul, M. R. 1991. "Lexical Choice: Towards Writing Problematic Word Lists." *International Review of Applied Linguistics in Language Teaching* 29 (1): 45–60.