LINGUISTIC DEVIATIONS IN THE WRITTEN ACADEMIC REGISTER OF DANISH UNIVERSITY STUDENTS

JONAS NYGAARD BLOM, MARIANNE RATHJE, BJARNE LE FEVRE JAKOBSEN, ALEXANDRA HOLSTING, KENNETH REINECKE HANSEN, JESPER TINGGAARD SVENDSEN, THIT WEDEL VILDHØJ AND ANNA VIBEKE LINDØ

ABSTRACT

Danish university students are often criticised for a general lack of proficiency in orthography, punctuation and grammar in the academic register. However, there has been limited empirical substantiation to support the claim. In this paper, we present the results of a study of linguistic deviations in university assignments written by first-year Journalism and Danish students at the University of Southern Denmark (N = 100 students). The results show that the majority of both groups struggles with Danish orthography and punctuation when writing academically, which seems to confirm some of the assertions made by the critics. However, it is argued that the inherent conflict of orthographic and punctuation principles in Danish as well as the specific characteristics and challenges of academic writing are more probable causes than the claimed general decline in the writing proficiency of students.

[1] INTRODUCTION

In recent years there has been a stream of criticism in the media concerning the allegedly poor writing skills of Danish university students. Most notably, the Dean of Roskilde University, Hanne Leth Andersen and the former Minister of Research and Education, Esben Lunde Larsen, have publically argued that the academic writing skills of students are more deficient than proficient (Borre 2014, Hjortdal 2014). External university examiners have also claimed that students do not master basic Danish orthography, punctuation and grammar in addition to other textual issues (Dahl 2013). One external examiner has even compared the writing and reasoning skills of university students to that of eleven-year-olds (Schultz-Jørgensen 2011). This criticism seems to be part of, or at least related to, a larger agenda concerning increased student intake at Dan-

ish universities (Danmarks Evalueringsinstitut 2015) and, as a consequence, a suspected decline in the writing capabilities and general academic competences of students (Bukajewitz & Husted 2015). However, while the negative perception in the media implies that the problem is wide-ranging, there is limited, if any, scholarly proof that this is actually the case (cf. Krogh 2009). Bearing in mind that proficiency in writing is considered a prerequisite for academic achievement and excellence in most studies in the humanities and social sciences, systematic empirical analyses are needed, firstly, to determine whether there is indeed a general problem and if so, its nature and scope, and secondly, to properly address and resolve the potential issues arising. The aim of this paper is to lay the foundation for such an approach by examining empirically which types of linguistic deviations, in terms of orthography, punctuation and grammar, university students make when writing academically, and by debating the reasons why such deviations may occur, based on the theory of orthographic principles as well as on the lexical and syntactic properties of the academic register, i.e. the functional text variety used for academic purposes that corresponds to the situational settings and language norms of academic institutions (cf. Halliday & Hasan 1985: 29). In doing so, we are contributing to the body of research into the normative aspects of language proficiency among pupils and students in the Danish education system, which up to now has mainly focused on primary, secondary and upper secondary schools. The paper is also part of a larger research project called Skriftsproglighed på Universitetet ('Written Language at the University'), which investigates further aspects of the language proficiency of Danish university students: for example, argumentation and adherence to academic writing conventions (Holsting et al. 2017).

Our study looks at the quantitative measurement and analysis of linguistic norm deviations in a corpus of written university assignments by first-year undergraduate students on the Journalism and Danish programmes (N = 100) at the University of Southern Denmark (SDU). We have chosen linguistic norm deviations as a topic because it is often highlighted as a central issue in the negative appraisal of university students, and because, according to the Danish university law, spelling and wording in bachelor's and master's theses and other types of larger-scale academic assignments must be assessed during examinations ('Bekendtgørelse om eksamen og censur ved universitetsuddannelser', § 26). Furthermore, Danish and Journalism students are taught and tested in spelling, grammar and adherence to official and academic language norms as part of the basic curriculum and learning objectives, since the students are learning to be professional text writers, and their job opportunities very much depend on their ability to write according to general norms. It must be

stressed, though, that other levels of discourse are equally important, or perhaps even more so, when assessing the students' writing proficiency within the academic register. For instance, in a related setting, Brink, Elbro and Johannsen (2014) have documented that in Danish high school papers, content influences grading to a higher degree than formal language deviations. However, the study also showed that certain types of orthographic deviations (comma deviations in particular) have a significant effect on grading, especially in low-graded papers. Consequently, we consider linguistic norm deviations an important, though by no means exclusive text feature that may affect the assessment of students' writing skills. The adherence to norms of spelling and punctuation are often treated differently than aspects more specific to the academic register. While the latter are acknowledged as challenges for new academics (cf. Biber 2006), the ability to spell correctly is taken for granted, whereby deviations are often referred to as the result of sloppiness, rather than lack of proficiency. In one of the most widely used handbooks on academic writing in a Danish context, the student is advised to be careful when putting the final touches to a text and to eliminate spelling, grammar and punctuation errors in order to prevent them from overshadowing the important part of the text (Rienecker & Stray Jørgensen 2005: 73). This implies that these errors can be identified and corrected by the student, if only he or she is careful enough.

In the present paper, we approach our topic by first giving a critical account of the challenges connected to Danish orthography and punctuation in the academic register and the principles that govern Danish orthography and punctuation. We then outline the design of the study and present the data. This is followed by a discussion of the methodological issues in the quantitative measurement of linguistic deviations. We then go on to describe our approach to the coding of the corpus. This in turn leads to our results, which we present, conclude upon and finally discuss in the closing section.

[2]   CHALLENGES CONCERNING ORTHOGRAPHY AND PUNCTUATION

Although they are part of the core syllabus in Danish primary, secondary and upper secondary schools, orthography and punctuation pose a challenge for many Danish adolescents (Undervisningsministeriet 2002, Johannsen 2012). This, however, is hardly a new trend. Indeed, a recent study by Jervelund and Schack (2016) shows that although 9th formers in secondary school in general have become slightly worse at spelling in tests compared to pupils in 1978, the orthography tests have also become more demanding. Accordingly, the recurrent spelling and punctuation issues perhaps suggest a challenging orthographic system rather than evidence of a new, linguistically incapable generation of

adolescents.

Danish orthography is, undeniably, very tricky, largely due to the frequently disproportionate relation between pronunciation and spelling (Juul & Sigurdsson 2004), the sometimes conflicting spelling principles (Hansen 1999, Jervelund 2007) and a complex set of comma rules that few people master fully (Jacobsen 2010: 1111).

In Danish, the basic principle of orthography is the phonematic principle: the notion of correspondence between letter and phoneme in standard Danish pronunciation (Hansen 1999, Jervelund 2007: 11). While this may hold true to a certain degree, modern spoken colloquial Danish is rapidly distancing itself from its written version as the result of widespread assimilation and reductions in Danish (Schachtenhaufen 2013) and a series of vowel changes (Grønnum 2005: 330 ff, Brink 2013). For instance, the central vowel in the word *sta**ti**stik* is prone to shift towards either [e] or [ə], resulting in the misspelling *statestik*, and also prone to reduction, leading to the misspelling *statstik.* Such incongruences between spelling and pronunciation illustrate that, in some cases, the phonematic principle applies to (hyper-)distinct pronunciation variants of individual lexemes in a particular perception of a chronolectal variant of standard Danish. Consequently, language users who are not used to pronouncing Danish in this way – for example because they are young and their chronolect differs from the official perception of canonical pronunciation in dictionaries – may be more prone to spelling deviations (Elbro, Bors & Klint 1998).

It may also be argued that the principle of writing words the way in which they were written in the past (the so-called tradition principle) must be considered significantly more dominant in Danish orthography than the phonematic principle. This point is also reflected in the official principles for determining Danish orthography applied by the Danish Language Council (Dansk Sprognævn, no date).

As a consequence, orthographic proficiency in Danish is very much dependent on a lexical knowledge of the way words are traditionally written according to the official norms, rather than only – or even primarily – being based on a logical correspondence between pronunciation and spelling. This is also governed by the somewhat contradictory application of the principle of language usage, which seems to make room for new spelling alternatives; however, only based on the writings of so-called "good and proficient" ('gode og sikre') language users (Kulturministeriet 1997: § 1.4.) The Danish Language Council gives

no precise definition[1] of what "good and proficient" entails, so it must be assumed that it applies to people who are already adhering to the norms. In other words, a change in norm must be based on the language use of people who already stick to the norm in most cases, sometimes resulting in circular argumentation and an inclination for keeping things as they are, instead of changing them according to modern innovations in spelling, by young people, for instance. A related issue is the divergence between pronunciation and the morphematic principle that aims to conserve the traditional inflectional/conjugational and derivative system in Danish spelling (Hansen 1999). This principle particularly influences nominal and verbal conjugations ending in (*r*)-*er*, leading to misspellings since the 'r' does not represent an individual phoneme. Thus, the infinitive *insistere* (to insist) is pronounced similarly to the present tense *insisterer* (insists) (i.e. [ensiˈsdeˀʌ]) and the pronunciation of the singular form *genre* ('genre') is identical to that of the plural form *genrer* ('genres') (i.e. [ˈɛaŋʁʌ]), both of which lead to spelling mistakes by pupils and students. We will return to this point in more detail at a later stage.

When it comes to punctuation, the challenges increase. This is due to several facts concerning the principles for the use of commas in Danish. First of all, Danish comma rules are rather strict and allow for very little individual variation. This means that variations are typically considered to be deviations or plain errors. This is the opposite of, say, the Swedish *tydlighetskomma* ('clarity comma'), where the use of commas is "rather free" (cf. *språkbruk.fi*). Secondly, at least one of the systems used in Danish – the one that is based on the so-called grammatical comma – requires many commas in texts with frequent subclauses, compared to other semantically-oriented comma systems. This means that the possibility of omitting commas and thereby making errors increases. In addition, few language users seem to master the grammatical rules in full, as postulated in, what we assume are, hyperbolic terms by Erik Hansen, the former Chairman of the Danish Language Council: "The grammatical comma is so ridiculous that we have to do something. There are only 25 people in Denmark who know how to place it properly" (Hansen 2000, Kristeligt Dagblad (our translation)). Finally, there are currently two comma systems in Danish, although, for political reasons, they are described as one system with freedom of choice with regard to placing a comma in front of non-parenthetical subclauses. The choice must, however, be consistent throughout the text.

Concerning the academic register, the inherent challenges of Danish or-

---

[1]   However, Diderichsen & Schack (2015) do attempt to define the "good and proficient" language user empirically. A text with many orthographic, morphological, and syntactic mistakes is also likely to lack textual coherence and style (Diderichsen & Schack 2015: 6).

thography and punctuation multiply caused by the general tendency for technical vocabulary and phrasal density (Freeman et al. 2017: 5) as well as specific syntactic characteristics. While technical vocabulary does not necessarily lead to words that are (very) difficult to spell, the use of technical terms in academia is often confined to highly specific content, resulting in a vocabulary that differs from general language usage and may even be entirely novel for newly started students. Granted that spelling is partially memory-dependent (Kreiner & Gough 1990), instances of low-frequent lexemes in academic vocabulary may result in spelling deviations. In addition, lexical density often leads to instances of nominal compounds, which in Danish are the main cause of one of the most frequent spelling issues (Jervelund 2007, Jervelund & Schack 2016, Heidemann Andersen & Diderichsen 2011). Finally, the syntactic characteristics of academic discourse may lead to punctuation deviations, although not necessarily caused by the frequency of subordinate complement and adverbial clauses, which is actually lower than in spoken registers according to Biber & Gray's large scale corpus study of academic discourse (2010: 8), but rather by the use and frequency of dense phrases and relative clauses, which is generally higher in academic registers than in spoken language (ibid.). For instance, when the initial phrase in a clause is dense and lengthy, language users tend to mark it with a comma in Danish. However, the comma rules dictate that the initial clausal constituent should not be suceeded by a comma, regardless of length and density, unless the constituent is a subclause or immediatly repeated by a pronoun. Furthermore, commas around relative clauses are particularly tricky when they are embedded in phrases that do not conclude the main clause. In such instances, and depending on the applied comma system, the language user needs to be able to discern between appositive (parenthical) and determinative clausal function and/or be able to locate clausal boundaries in front of and behind the relative clause in order to punctuate according to the rules.

In sum, we expect that spelling and punctuation are still a challenge at university level, especially for new students.

[3] DESIGN

We opted for a case study of first-year students of Danish and Journalism at the University of Southern Denmark. These groups are particularly suitable for the present study because they are expected to be fluent and well-nigh flawless in written Danish. Furthermore, the two groups are granted admission in different ways. At the time of the study (2015), Danish students were granted admission on the basis of an average grade, although in reality all applicants were admitted. All Journalism students, on the other hand, take an entrance

exam, which includes orthography, punctuation and grammar tests that influence the final grades and, in turn, the probability of admission. We expect that this variable has a positive influence on the students' orthographic and grammatical performance, considering that the students who are admitted have not only previously been tested in, but also primed towards orthography, punctuation and grammar as significant learning objectives at the university.

The study was designed as an experiment in which the students were asked to write an academic assignment at the University under supervised conditions. The experimental approach ensured that the textual performance of the students was comparable and not influenced by other parties. In this regard, it may be assumed that regular home assignments may have led to tainted data due to fellow students, friends, partners or relatives aiding the student by proofreading.

The assignment was constructed as a basic linguistic text analysis and consisted of a short extract on linguistic reference theory written for high school and university students, which the students were asked to account for and apply in an analysis and evaluation of a manipulated newspaper article. Thus, the students were tested in basic academic skills in ascending taxonomic levels (Biggs & Tang 2007). In addition, the theory text included four technical linguistic terms that were expected to be more or less novel for most of the students: *kohæsion* (cohesion), *proform* (pro-form), *anaforisk* (anaphoric) and *kataforisk* (cataphoric). Since, like all exams and assignments at SDU, the test was conducted on a computer, spelling and grammar controls in word processing programmes were allowed, and the students were neither prohibited from using nor explicitly encouraged to use dictionaries in digital or physical form. This was done in order to mimic the regular conditions for assignments at the University. The time limit for the assignment was set at one hour based on a pre-test that indicated sufficient time for proofreading approximately one normal page, which was the recommended size of the assignment.

As listed in Table 1 below, 88 Journalism students and 72 Danish students took part in the test. However, the gender distribution proved to be askew among the participating Danish students, which resulted in an imbalanced data set. Few males study Danish at the University of Southern Denmark, and even fewer chose to participate in the experiment (8 in total). In order to compensate for this bias in gender, undersampling was used. This led to a sample of 50 texts by the Journalism students and 50 texts by the Danish students, of which 16% were done by male students equivalent to the distribution of male, first-year Danish students at SDU (16 out of 97 students, class of 2015). Sampling was done randomly with the exception of the above-mentioned males, two students

who were excluded due to plagiarism (i.e. direct copy-paste of sections from the theory text), and one suspected case of severe dyslexia.

	Journalism	Danish
Participants in test	88	72
Sample texts in corpus	50	50
Gender in sample	24 ♀ (48%), 26 ♂ (52%)	42 ♀ (84%), 8 ♂ (16%)
Age in sample	22.2 (mean)	23.4 (mean)
Words per sample text (mean)	452 words	446 words
N (sample)	22,310 words	22,647 words

TABLE 1: Sample texts and participant distribution.

[4]   DEFINITIONS, CATEGORISATIONS AND CODING

Linguistic deviations are difficult to define and delimit, even harder to categorise and quantify (cf. Jervelund 2007: 35), and impracticable to compare across studies that examine different texts in different genres in different settings by different types of language users. This is due to a number of factors.

Firstly, the concept of *deviation* presupposes well-established language norms. However, with the exception of entries, rules and descriptions in the official orthography dictionary, *Retskrivningsordbogen*, there is no common scholarly consensus on what constitutes a linguistic deviation. A case in point involves the so-called pleonastic conjunctions in Danish such as *fordi at* ('because that') and *hvis at* ('if that'), which are considered errors by conservative language users, but which are highly common and which some linguists argue are grammatically motivated rather than just being superfluous (Hansen 1975). Another example is the oblique case in Danish, which conservative language users argue cannot be used, when pronouns function as head of NPs in subjects: e.g. *Dem, der hvisker, lyver* ('Those who whisper are lying'). This normative opinion, however, has been questioned and rejected on grammatical grounds by the likes of Hansen and Heltoft (2011: 29-30). Consequently, the coding of linguistic deviations necessitates a transparent account of the normative discourse(s) on which the coding is based.

Secondly, deviations can be counted in a variety of ways. A trivial example is whether or not to count only unique deviations or to include repeated deviations. For instance, a student might use the spelling deviation *intereseret* (for *interesseret* ('interested')) ten times in a paper. If this is counted as ten deviations, rather than one, it may give the statistical impression that the student makes a lot of deviations, which would not necessarily have been the case if he

or she, by chance, had only used the word once, and thus could only deviate from the 'proper' spelling once. On the other hand, a critical reader (e.g. an external examiner) might perceive repeated deviations as more salient than singular deviations. Again, transparency is key, since it can be argued that both approaches, i.e. counting with and without deviation repetitions, are valid.

Thirdly, some types of words, phrases and syntactic constructions are more prone to deviations than others, and specific genres, registers and content often govern their occurrences. For example, appositions such as *Danmarks statsminister, Lars Løkke Rasmussen* ('The Prime Minister of Denmark, Lars Løkke Rasmussen'), are prone to comma deviations in Danish because the writer has to decide whether or not the apposition is parenthetical in order to punctuate properly. While appositions are presumably a relatively rare occurrence in many genres and registers, they are highly common in news journalism because journalists need to introduce the titles, occupations and names of their sources. Consequently, news journalists who are not familiar with the rules of appositional commas have a higher risk of making this particular deviation than writers who use appositions more rarely in other genres and registers. In other words, quantified deviations are not directly comparable across different genres, registers and content unless the likelihood of given deviations in specific contexts is taken into account. To our knowledge, this approach is yet to be adopted and presupposes large scale corpus analysis across a multitude of genres and registers. While this is not the purpose of the present article, we choose to approach the quantification of deviations with reservations.

In this article, we consider linguistic deviations in written discourse as formal divergences from the orthographic and grammatical norms as well as lexical conventions within a given societal or institutional language community, in our case Danish universities. Furthermore, we delimit the range of linguistic deviations in this study so that it only covers the least disputable types. We consider these to be formal divergences from: 1) the orthographic rules, the rules of punctuation, entry words, examples and explanations in the official Danish orthography dictionary, *Retskrivningsordbogen*; 2) the idiomatic word entries, including expanded predicates, in the dictionary *Den Danske Ordbog*; 3) the grammatical conventions for congruency, syntax and cohesion in standard Danish as described in *Grammatik over det Danske Sprog* (Hansen and Heltoft 2011) and normative reference works such as *Håndbog i Nudansk* (Jacobsen and Jørgensen 2013); and 4) the conventions of space between words and lack of space between punctuation and words. Additionally, indisputable wrong, missing or superfluous word(s) and anacoluthon are considered semantic deviations based on the coder's knowledge of Danish semantics. In order to ensure full

transparency, we have added an appendix with descriptions of our coded deviation categories. In our counting we have chosen to include repetitions of deviations to reflect the presumed salience of repeated deviations from the reader's perspective.

In addition to this approach, we have applied an intercoder reliability test to our coding. The main author of the present paper initially coded all deviations, including repeated deviations, in the entire sample combined with an automated search for orthographic and grammatical deviations conducted by the constraint grammar application *DanProof*[2], which is coded according to the rules and entries of *Retskrivningsordbogen* (*RO*) and the constraints of Danish grammar and syntax (Bick 2015). It should be noted that DanProof does not supply a fully reliable coding, and the application has simply been used to crosscheck for coding mistakes. In addition to the main coding, the second author has followed the same coding procedure in an excerpt of the corpus (10%). The applied intercoder test had a reliability score of 0.856 (Krippendorffs α), which can be considered acceptable. The primary disagreement was caused by the coding of quotations (*RO*, § 59).

Taking the two comma systems in Danish into account, we chose to code all misplaced, missing and properly placed commas according to both comma systems as they are formulated in *Retskrivningsordbogen*. Afterwards, we determined the applied comma system by comparing the distribution of incorrect commas in both variants. The one with the least deviations compared to the properly placed commas was chosen as the default in the specific text. Accordingly, we coded the commas from the perspective of a reader who knows both systems and will assume that the one with the least deviations from the rules is the one applied. Our approach shows that the majority of the students in our study, more specifically 45 Journalism students (90%) and 46 Danish students (92%), place commas in front of non-parenthetical subclauses - an option which is *not* recommended by the Danish Language Council, but which is nevertheless a popular choice for many: e.g. the Danish media.

[5]   RESULTS

In this section, we start by presenting the general results before focusing on the most frequent types of deviations. As seen in Table 2, the Danish students made an average of 64 deviations per 1,000 words, while the Journalism

---

[2]   DanProof is an IT-based pedagogical spelling and grammar checking system for Danish, developed by Eckhard Bick as part of the VISL research programme at the University of Southern Denmark (http://visl.sdu.dk/visl/da/). The DanProof software programme checks a variety of errors at word level such as orthography, inflection, word order, and removal and insertions of words, using the Constraint Grammar formalism.

students did significantly better with 47 deviations. Repeated deviations are included in this account, and the significance level has been calculated as a two tailed T-test with the deviations by the Danish and Journalism students as two independent variables (t-value = 3.0034, p-value = .003388). As expected, the result indicates that the orthographic entrance exam has a positive influence on performance, although other variables might also play a role.

	Journalism	Danish
Deviations/1,000 words (mean)	47	64
Standard deviation	27.5	31
Minimum number of deviations/1,000 words	7.2	7.8
Maximum number of deviations/1,000 words	163.1	145.1

TABLE 2: Deviations in corpus

The standard deviation and the difference between the minimum and maximum numbers of deviations show that there are noteworthy differences between the students. Some have serious difficulties, while others fare much better. However, no student in the sample was capable of producing 450 words (approx. 1 normal page) without a minimum of 3 deviations. Now, the key question is whether or not an average frequency of 47 and 64 deviations per 1,000 words is a lot, a few or somewhere in between. Compared to other studies, it seems like a lot. For instance, Andersen (1992) and Johannsen (2012) found fewer deviations (15.6 and 21.1 deviations per 1,000 words) in their studies of deviations in exams and papers in Danish upper secondary schools. In comparison, the University students in question seem to be doing significantly worse, which seems to substantiate rather than counter the criticism of the students' lack of writing proficiency in this particular case. However, such comparisons are hampered by the aforementioned sources of errors including incomparable genres, settings, language users and conditions under which the texts were produced, as well as by differences in definitions and coding of deviations. Consequently, the comparisons should be interpreted very cautiously. However, it should be safe to assume that in general the students are not showing a high degree of proficiency in following standard writing norms.

In line with the findings of Brink, Elbro and Johannsen (2014), Johanssen (2012: 19) and Rathje (2013: 344), Table 3 shows that punctuation, and commas in particular, are by far the most frequent deviation, followed by different types of spelling issues.

	Occurrences	Percentages
Punctuation	1,220	49.6%
Spelling	708	28.8%
Semantics	207	8.4%
Reference	102	4.1%
Layout	75	3%
Concord	58	2.4%
Idioms	55	2.2%
Syntax	37	1.5%
Total	**2,462**	**100%**

TABLE 3: Occurrences of deviation types in sample

[5.1] *Commas*

On average, the Danish students deviate in approximately every third comma (35%), while the Journalism students deviate in every fourth (25%). Furthermore, the comma deviations are distributed in such a way that 34% of the Journalism students have a low amount of comma deviations (<10 per 1,000 words) compared to 12% of the Danish students.

As shown in Table 4 below, the missing commas are primarily lacking before and after subclauses (e.g. *I den sammenhæng [,] som det står skrevet i artiklen [,] giver sætningen lige pludselig mening* ('In the context, in which it is written in the article, the sentence suddenly makes sense')). It should be noted, though, that the difference in frequency between initial and final comma is primarily caused by the general tendency of subclauses to be concluded by a full stop instead of a comma. This leads to fewer commas after subclauses, and thus fewer potential comma deviations, compared to commas before subclauses.

In addition, some of the texts are characterised by complex clausal structures which generate a series of potential comma issues as well as syntactic deviations, e.g.: *Det vurderes ikke, at grunden er læserens mangel på viden indenfor et bestemt område, og derfor umuligt kan gætte sig til hvad der refereres til, da vi må gå udfra, at alle er klar over, hvad både et komma, en undersøgelse, og en debat er.* ('It is not estimated that the reason is the reader's lack of knowledge within a given area and therefore cannot possibly [anacoluthon] guess at what is being refered to since we must assume that everone is aware what a comma, a study, and a debate is.'). Although the use of multiple complement and adverbial clauses does not necessarily resonate with the typical syntactic characteristics of academic registers, at least in an English context (Biber and Gray 2010), examples such as these do occur with some regularity in the corpus, perhaps in-

dicating that some of the students are not yet familiar with the academic writing norm. Regardless, the use and frequency of subclauses must be considered a main reason for punctuation deviations.

Position	Occurrences	Texts with deviation
In front of subclauses ('startkomma')	71.1% (409)	77
Behind subclauses ('slutkomma')	15.1% (87)	44
Between coordinated main clauses	9.7% (56)	32
In front of and behind independent phrases	3.5% (20)	11
Between coordinated phrases (serial comma)	0.3% (2)	2
Between main clause and VP-ellipsis	0.2% (1)	1

TABLE 4: Missing commas

In comparison, the misplaced commas are primarily found after initial phrases in clauses (e.g. *I linje 11, ser man hvordan der peges tilbage* ('In line 11, you see how it refers back')) and before infinitives (e.g. *Derudover kan proformer bruges til, at referer [sic] til en hel sætning* ('Additionally, pro-forms are used to refer to an entire clause')), as seen in Table 5.

Position	Occurrences	Texts with deviation
After initial phrase in main clause or subclause	25.5% (105)	49
Before infinitives	15.8% (65)	40
Mix-up of comma and full stop	15.8% (65)	26
After *hv*-conjunction phrase	7.8% (32)	22
In front of coordinating conjunction with no following main clause or subclause	6.6% (27)	17
Between main clause and subject-ellipsis	6.3% (26)	21
After preposition or adverb initiating subclause	2.9% (12)	9
In front of *som* (not subordinating clause)	2.7% (11)	9
In front of and after non-parenthetical apposition	2.2% (9)	3
Other misplacements	14.4% (59)	38

TABLE 5: Misplaced commas

The most frequently misplaced comma is after the initial phrase in main clauses and subclauses, in some cases caused by the students' tendency to use dense constituents before the finite verb, e.g.: *Efter introduktionen til Dansk*

Sprognævns undersøgelse, får man at vide at den kommer på baggrund at et bestyrelsesmøde. ('After the introduction to the study by the Danish Language Council, we are told that it is based on an editorial board meeting.'). This is an example of how the phrasal density of the academic register influences punctuation. By compressing two nominal predicates (*introduktionen* and *undersøgelse*) into the initial constiuent, the inclination to insert a comma, as a kind of constituent boundary, rises.

[5.2] *Spelling issues*

The most noteworthy and recurrent spelling issue in the data set is the misspelling of the present-tensed verb *refererer* ('refer/refers'), which is very frequent in the texts on account of the assignment topic: reference theory. The verb is misspelled as either *referer* (imperative mood) or *referere* (infinitive) in 48 of the texts, 141 times in total. Compared to the total number of 708 spelling deviations in the corpus, this particular deviation alone amounts to one fifth of all the spelling deviations.

The misspellings of *refererer* may be caused by grammatical confusion. However, it is more likely that phonetic equivalence and/or structural parallelism are causing the problems. The verb *referere(r)*, pronounced as [ʁɛfəˈʁɛʔʌ] is phonetically identical in both the present tense and the infinitive form, making it impossible to discern the spelling based on pronunciation in spoken discourse. This has been a problem since the 19th century, but is supposedly a growing problem because of increasing reductions in Danish pronunciation (Detlef & Lund 1986). In addition, the verb consists of three identical parallel structures (*er-er-er*) in the present tense. As observed by Lund (1985), such structures are often very difficult to segregate for the eye and are thus prone to spelling deviations. Cf. also Jervelund & Schack (2016: 35) where *refererer* was the second most problematic word in a spelling test.

Another significant issue is the misspelling of technical academic words, which were introduced in the theory text, and which the students presumably did not know beforehand. *Proform* is such a word, in Danish: *en proform* ('a proform') or *proformen* ('pro-form-*the*'). The particular word is misspelled in 35 of the texts, a total of 87 times, and provides several variants in the corpus: *et proformen*, *et proforme*, *en proforme*, *proformet*, *proformeren*, *proformere*, *et proform*, *profem*, *profemer*, *profemerne*, *profomer*. The linguistic terms *katafor* ('cataphora') and *anafor* ('anaphora') are also misspelled in adjectival form as for instance *karaforisk, kataforsik, kateforisk*. These results confirm that technical vocabulary in academic registers is subject to spelling deviations among students that are not (that) familiar with the terms beforehand.

The data also shows a variety of spelling issues, although some seem to be caused by typing errors rather than lack of linguistic competence: for example, *atiklen* for *artiklen*, *elemeter* for *elementer, solget* for *solgte*, and *sporlige* for *sproglige.* The last two spelling deviations are examples of the phenomenon *reversalfejl* ('reversal error'), in which two letters are reversed (Detlef & Lund 1986). Others may be considered deviations caused by lack of competence. It is, of course, difficult to distinguish between typos and lack of competence (cf. Rathje 2013 for a discussion). In this group, we find different categories of spelling deviations. The categories are based on Noesgaard (1945) and Detlef & Lund (1985):

(i) Silent letters that are missing (e.g. *mangelfulhed* (for *mangelfuldhed*), *bevist* (for *bevidst*), and *forman* (for *formand*))

(ii) Double/single consonants (e.g. *ellement* (for *element*), *verballet* (for *verbalet*), *kolega* (for *kollega*), and *resonere* (for *ræsonnere*))

(iii) Consonant mix-up (e.g. *sjældend* (for *sjældent*), and *af vide* (for *at vide*))

(iv) Vowel mix-up (e.g. *rattet mod* (for *rettet mod*), *for* (for *får*), and *irreterrer* (for *irriterer*))

(v) Word mix-up (e.g. *uden at set* (for *uanset*), and *faldgruppe* (for *faldgrube*))

(vi) Lack of the letter *r* in present-tense verbs or excess *r* in infinitives (e.g. *fungere* (for *fungerer*), *citere* (for *citerer*), and *at hører* (for *at høre*))

(vii) Separate words in compounds (e.g. *sprog brug* (for *sprogbrug*), and *uddannelses institutioner* (for *uddannelsesinstitutioner*))

(viii) Inflection deviations (e.g. *syntes* (for *synes*))

(ix) Annulment of vowel (e.g. *tydligt* (for *tydeligt*) and *forstålig* (for *forståelig*))

(x) Annulment of consonant (e.g. *forsætte* (for *fortsætte*), and *nød til* (for *nødt til*))

(xi) Addition of vowel: (e.g. *forståes* (for *forstås*))

(xii) Capital letters (e.g. *dansk sprognævn* (for *Dansk Sprognævn*))

(xiii) Foreign words (e.g. *marathon* (for *maraton*)).

Among the most frequent spelling deviations in the last year of Danish primary school and in high school are separate words in compounds and problems with the letter *r*, especially in present-tense verbs and infinitives (Jervelund & Schack 2016, Undervisningsministeriet 2002). This accords in part with our results. As illustrated above, issues with the letter *r* are prominent in the corpus, though primarily due to the misspelling of the verb *referere*. Separate words in compounds are also present at university level, although the frequency is not strikingly high (32 in total, occurring in 25% of the texts) in spite of the lexical density of the academic register.

More noteworthy, some of the students struggle with silent letters and consonant doubling, which are traditionally the categories with which the lowest graded pupils in primary school struggle. For example, a study showed that 8th and 9th formers regard these spelling deviations as severe mistakes because, as they say in an interview: 'We have learned this since nursery class' (Kristiansen & Rathje 2014).

[6]   CONCLUSION AND DISCUSSION

Overall, our study shows that punctuation is a challenge for most of the newly started Danish and Journalism students, and that some of them struggle with basic spelling issues as well, such as silent letters and consonant doubling. However, the relatively high frequency of deviations is mainly caused by specific issues, in part relating to the properties of the academic register. These issues concern complex spelling patterns such as structural parallelism, technical academic vocabulary and comma deviations caused by phrasal density, as well as the frequent use of subclauses.

In addition, the methodological challenges of quantifying deviations make it difficult fully to validate a negative appraisal. Moreover, we prefer to be cautious concerning the generalisability of our results. The students who participated in the experiment are from the University of Southern Denmark (SDU), a provincial university, which attracts other types of students than the University of Copenhagen and Aarhus University. Among the Danish universities, SDU and Aalborg University have the highest number of admitted students at undergraduate level with *low* average high school grades (Danish Evaluation Institute 2015)[3]. Therefore, it must be expected that some, or even many, have moderate or poor writing skills when commencing a university course.

The second main result shows that the entrance-examined Journalism stu-

---

[3]   This does not mean that the majority of the students at SDU performed less than average, though.

dents fared better than the students of Danish. Consequently, testing new students in grammar, orthography and punctuation as part of an entrance exam at the university is a tempting, although also debatable, approach to the problem. Not only are such exams costly, they also carry an implicit criticism of the learning objectives and outcomes of the Danish school system. In other words, universities are reconsidering what the students (should) have already proven they have learned in their previous preparatory education. A more constructive alternative would be to seek a better alignment of learning objectives at the various levels of the Danish educational system. In this respect, the results of our study outline the specific deviation types that should be addressed and focused on, provided that there is indeed a need and a wish for reducing the number of linguistic deviations. However, the last premise is debatable, especially considering the above-mentioned incongruences and complexities of the Danish orthographic system and punctuation rules, which seem to engender deviations, almost regardless of writing proficiency. However, the alternative solution, to simplify the principles and rules and to accept more spelling variations, has not yet proven to be popular among the Danish population, though this would potentially solve the issue once and for all.

REFERENCES

Andersen, Torben. 1992. *Stavefejl. En undersøgelse af stavefejlsmønstre i gymnasiet.* Undervisningsministeriet.

Biber, Douglas. 2006. *University Language. A corpus-based study of spoken and written registers.* Philadelphia, NL: John Benjamins.

Biber, Douglas & Bethany Gray. 2010. Challenging stereotypes about academic writing: Complexity, Elaboration, Explicitness. *Journal of English for Academic Purposes 8*, 2-20.

Bick, Eckhard. 2015. DanProof: Pedagogical Spell and Grammar Checking for Danish. International Conference: Recent Advances in Natural Language Proceedings, 55-62.

Biggs, John & Catherine Tang. 2007. *Teaching for Qualitative Learning at Universities.* 3rd edition, Open University Press.

Borre, Martin. 2014. Politisk flertal støtter opgør med undervisningen på universiteterne. *Politiken.dk.* 10.4.2014.

Brink, Lars. 2013. Nyere danske lydlove. Udtaleændringer siden 1955-generationen. *Mål & Mæle*, vol. 36. , 1, Copenhagen.

Brink, Eva, Carsten Elbro & Stine Johannsen. 2014. Sproglige fejl og karakterer i danske gymnasiestile. *Mål & Mæle* 2, 5-9.

Bujakewitz, Tina Michler & Tinemarie Husted. 2015. *Er kvaliteten fulgt med optaget?* Master Thesis. Aalborg University.

Dahl, Henrik. 2013. Den massehumanistiske trosbekendelse. *Kommunikationsforum.dk.* 25.6.2013.

Danmarks Evalueringsinstitut. 2015. *Fordelingen af det stigende optag på universiteterne*, Copenhagen.

Dansk Sprognævn. 2012. *Retskrivningsordbogen*, 4th edition, Dansk Sprognævn.

Dansk Sprognævn (no date) *Hovedprincipper for fastlæggelse af retskrivningen*, https://dsn.dk/om-os/sprognaevnets-opgaver/sprognaevnet-og-retskrivningen-1/hovedprincipper-for-fastlaeggelse-af-retskrivningen, Dansk Sprognævn.

Undervisningsministeriet. 2002. *Hvordan staver studenterne?*, Copenhagen.

Det Danske Sprog og Litteraturselskab. 2017. *Den Danske Ordbog. Online version*, http://ordnet.dk/ddo.

Detlef, Claus & Jørn Lund. 1986. *Stavning.* Dansklærerforeningen.

Diderichsen, Philip & Jørgen Schack. 2015. Jagten på den gode og sikre sprogbruger. *Nyt fra Sprognævnet* 2015/3. 1-8.

Elbro, Carsten, Ina Bors & Dorthe Klint. 1998. Predicting dyslexia from kinder-garten. The importance of distinctness of phonological representations of lexical items. *Reading Research Quarterly* 33(1), 36-60.

Freeman, David E., Yvonne S. Freeman & Ivannia Soto. 2017. *Academic Language Mastery: Grammar and syntax in context*, Corwin (Sage).

Grønnum, Nina. 2005. *Fonetik og fonologi.* 3rd edition, Akademisk forlag.

Halliday, M. A. K. & Ruqaiya Hasan. 1985. *Language, Context, and Text: Aspects of Language in a Social-Semiotic Perspective.* Oxford/Geelong: Deakin University Press.

Hansen, Erik. 1975. At eller ikke at. *Nyt fra Sprognævnet* 14, 1-4.

Hansen, Erik. 1999. *Skrift, stavning og retstavning.* 3rd edition, Hans Reitzel.

Hansen, Erik & Lars Heltoft. 2011. *Grammatik over det Danske Sprog.* Bind I. Syddansk Universitetsforlag.

Heidemann Andersen, Margrethe & Philip Diderichsen. 2011. Sjældne og sære sammensætninger. *NyS – Nydanske Sprogstudier 41*, 40-65.

Hjortdal, Marie. 2014. RUC's nye rektor: Undervisningen skal vendes helt på hovedet. *Politiken.dk.* 10.4.2014.

Holsting, Alexandra, Jesper Tinggaard Svendsen, Anna Vibeke Lindø, Ditte Lund Iversen, Jonas Nygaard Blom, Marianne Rathje, Kenneth Reinecke Hansen & Bjarne le Fevre Jakobsen. 2017. "Det første begreb i teksten er…" En undersøgelse af henvisningsformer i nyindskrevne universitetsstuderendes akademiske opgaver. *Rask 46*, 21-62.

Jacobsen, Henrik Galberg. 2010. *Ret og Skrift. Officiel dansk retskrivning 1739-2005.* Syddansk Universitetsforlag.

Jacobsen, Henrik Galberg & Peter Stray Jørgensen. 2013. *Håndbog i Nudansk*, 6. Ed. Politikens Forlag.

Jervelund, Anita Ågerup. 2007. *Sådan staver vi.* Dansklærerforeningens Forlag.

Jervelund, Anita Ågerup & Jørgen Schack. 2016. *Den der skriver d i gjort.* Dansk Sprognævn.

Johannsen, Stine. 2012. Fejl og karaktergivning i danske gymnasiestile. *Mål & Mæle* 1, 18-23.

Juul, Holger & Baldur Sigurðsson. 2004. Ortografien som handicap? Stavetilegnelse i Rødovre og Reykjavík. *Nys – Nydanske Sprogstudier* 32, 67-88.

Kreiner, David S. & Philip B. Gough. 1990. 'Two ideas about spelling: Rules and word-specific memory'. *Journal of memory and language, vol. 29 (1),*103-118.

Kristiansen, Anna & Marianne Rathje. 2014. 'Det kommer an på hvilken stavefejl det er': Unges holdninger til stavefejl i nye medier. *Nys – Nydanske Sprogstudier* 46, 103-131.

Krogh, Ellen. 2009. Faglighed og skriftlighed. *Gym Pæd 2.0* (2). Syddansk Universitet.

Kulturstyrelsen. 1997. '*Bekendtgørelse om Dansk Sprognævns virksomhed og sammensætning',* Copenhagen,

https://www.retsinformation.dk/Forms/R0710.aspx?id=123673

Lund, Jørn. 1985. Egernerne, søskendene og de mangehovedede uhyrer. Om strukturparallelisme. *Nyt fra Sprognævnet* 2, 1-3.

Noesgaard, Aksel. 1945. *Fejltyper i dansk Retskrivning*. København.

Rathje, Marianne. 2013. 'sku allgevel ha købt penis forlænger creme :)'. Afvigelser fra retskrivningsnormen i kommentarfeltet på *ekstrabladet.dk*. Inger Schoonderbeek Hansen, Tina Thode Hougaard & Peter Widell (Eds.), *14. Møde om Udforskningen af Dansk Sprog*, 333-353.

Rienecker, Lotte & Peter Stray Jørgensen. 2005. *Den gode opgave – håndbog i opgaveskrivning på videregående uddannelser*, 3rd edition, Copenhagen: Forlaget Samfundslitteratur.

Schachtenhaufen, Ruben. 2013. Fonetisk reduktion i dansk. Ph.d. Serie 15. Copenhagen Business School. Copenhagen.

Schultz-Jørgensen, K. (2011) 'Studerende skriver og ræsonnerer som 11-årige', *Politiken.dk*, 22.2.2011.

APPENDIX: LIST OF CODED DEVIATION CATEGORIES IN CORPUS

Category	Type
Layout	Lack of *or* excess space between words and between punctuation and words.
Spelling	*Letters*: Deviations from standard use of small and capital letters in sentence onset, proper names, titles, addressing pronouns and abbreviations (RO § 11-14). *Lexemes*: Deviations from entry words, examples and isolated semantic annotations in *Retskrivningsordbogen*, including certain word pairs with similar or identical pronunciation but different use, such as *ligge/lægge*, *ad/af*, *nogen/nogle*. *Morphemes*: Deviations from compounding, grammatical gender, standard conjugation and inflexion as determined in extensions to entry words and the orthography rules (§ 18-39) in *Retskrivningsordbogen*.
Punctuation	Deviations from punctuation rules (RO § 40-60), including full stop, semicolon, colon, comma, question and exclamation marks, dash, hyphen, parenthesis, quotation marks, and, in addition, apostrophe and acute accent (RO § 5-6).
Concord	Lack of grammatical concord/agreement.
Syntax	Deviations from common Danish word order in written registers; sentence/clause fragments.
Reference	Non-cohesive use of definiteness in nominals (given, new, generic information); endophoric reference with no referent, ambiguous referents or lack of concord across sentence boundaries.
Idioms	Deviations from idiomatic phrases, including contaminations, as listed in *Den Danske Ordbog*, comprising specification of predicate frames, i.e. predicate + (prepositional) argument positions (e.g. 'someone gives someone something', 'someone gives something (to someone)').
Semantics	Wrong, missing or superfluous word(s), anacoluthon.

CONTACTS

Jonas Blom
University of Southern Denmark,
blom@journalism.sdu.dk

Marianne Rathje
University of Southern Denmark,
Rathje@sdu.dk

Bjarne le Fevre Jakobsen
University of Southern Denmark,
Bjarne@sdu.dk

Kenneth Reinecke Hansen
Metropolitan University College
keha@phmetropol.dk

Jesper Tinggaard Svendsen
University of Southern Denmark,
Tinggaard@sdu.dk

Alexandra Holsting
University of Southern Denmark,
aho@sdu.dk

Thit Wedel Vildhøj
thitsko@gmail.com

Anna Vibeke Lindø
University of Southern Denmark,
avl@sdu.dk