Practice Report:
A blended learning approach to teaching NLP for a DH public

Gertrud Faaß
University of Hildesheim
Institute for Information Science
and Natural Language Processing
gertrud.faass/

Ulrich Heid
University of Hildesheim
Institute for Information Science
and Natural Language Processing
ulrich.heid@uni-hildesheim.de

Abstract

This paper reports about current practice in a staged approach to the introduction of NLP principles and techniques for students of information science (IIM) and of international communication and translation (ICT) as part of their curricula. As most of these students are rather not familiar with computer science or, in the case of IIM students, linguistics, we see them as comparable with students of the humanities. We follow a blended learning strategy with lectures, online materials, tutorials, and screencasts. In the first two terms, we focus on linguistics and its formalisation, NLP tools and applications are then introduced from the third term on. The lectures are combined with tutorials and - since the summer term 2017 - with a set of screencasts.

1 Background

The Natural Language Processing (NLP) BA course programme at the University of Hildesheim is part of the curriculum of students of International Information Management (henceforth IIM, 1st year (2nd term)) and of the BA in International Communication and Translation (ICT, 2nd year (1st term)). Additionally, students of other fields of the Humanities, like, for example, Political Science or Business Administration may choose the courses as part of their minor subject.

One may assume that *International Information Management* is a field of study usually not included in the Humanities; however, the IIM curriculum focuses mainly on methods of social sciences and empirical studies (such as, e.g., usability studies or qualitative surveys). The students of IIM are prepared for consulting jobs in industry and only to a much lesser extent for information retrieval, which relates more closely to NLP. The ICT students usually plan to work as professional translators, some specialize in the translation of technical texts. We find our BA students being less informed about formal modelling, which we however consider a key issue in introducing NLP methods and applications. Therefore, we decided to deal with this challenge in several didactic steps: General introduction, formalisation, and application.

Attempting the Blended Learning approach, we provide a mix of didactic devices: Face-to-face (FtF) oral presentations (lectures and tutorials, some accompanied by slides) are complemented by written scripts and background literature provided on an online platform (moodle), and - since summer 2017 - downloadable mp4 files containing audio recordings describing additional slides. Attending the lectures and tutorials cannot be made mandatory for the students, however, we intensively encourage a regular attendance in the FtF sessions, following the experiences of Dickson and Stephens (2016, p. 7) finding that "the most significant predictor of mark [. . .] was lecture attendance".

We have not found a similar programme at a German university, however, as (Neumeier, 2005, p. 164) rightly states, one has to decide for a combination of modes given the particular conditions at hand.

This practice report begins in section 2 with a description of the course programme and its main addressees. We then focus in section 3 on contents and methods utilized in our teaching and report on a first evaluation summarizing our experiences in section 4. Finally, in section 5, we summarize our programme and widen the scope, as we plan to offer our programme - or an adapted version thereof - to other potential addressees.

2 Course programme / addressees

The current curriculum concerning (computational) linguistics / NLP expands over 4 terms of 12-13

weeks each, addressing mainly two groups of students. IIM (about 120) and ICT students (less than 200) are to visit a general lecture that provides an *Introduction to Linguistics* ("Einführung in die Sprachwissenschaft") in their first term. This lecture is offered by the colleagues of the ICT team.

The ICT students afterwards focus on translation studies before returning to linguistics later. About half of them choose to focus rather on computational linguistics than on the translation of technical texts and attend the lecture called *(Formal) Description of Language for the Purpose of NLP* ("Sprachbeschreibung für die Sprachtechnologie (SBST)") in their 4th term. In subsequent terms, these students also attend one of the seminars offered by our team on *Machine Translation*, *Electronic Lexicography*, *Wordnets* or similar.

For students of IIM, the SBST lecture is mandatory in their second term. From their third term on, they can choose to focus their studies either on NLP or on Information Retrieval. Hence, only interested IIM students (about half of all) then continue with our lecture *Introduction to Computational Linguistics* ("Einführung in die Maschinelle Sprachverarbeitung"), though some also take courses on Information Retrieval or Man Machine Interaction. Lastly, seminars on *Corpus Linguistics*, *Corpus Processing Tools*, *Wordnets*, *Electronic Dictionaries* and similar are offered to them, supplemented by practical courses on scripting languages like *perl*, *python* or, again, *Corpus Based Analyses*.

3 Methods

From the perspective of Blended Learning (BL), we follow the definition of BL as "a combination of face-to-face (FtF) and computer assisted learning (CAL) in a single teaching and learning environment", based rather on practice than on research (Neumeier, 2005, p. 164). We do not plan for a (partial) replacement of FtF methods like O'Connor et. al (2011), as we rather attempt to achieve that "the on-line component becomes a natural extension of traditional classroom learning" (OConnor, 2011, p. 64). The courses are all organized "Module by Module", as described by Lisetskyi (, p. 34).

3.1 Lectures

Lecture (1) provides a general introduction to the history of linguistics, basic functions of language, and models of communication processes, followed by introductions to the basic linguistic disciplines,

each related to questions of applied linguistics with a focus on translation and international communication. The lecture is based on (Graefen, 2012). Students are to pass a written examination.

Lecture (2) is a three-part lecture describing

- Syntax (Constituency (A), Dependency including verb valency (B), Topological fields (C));

- Morphology (Inflection, Word Formation);

- Semantics (Lexical Semantics (including Word Nets), Compositional and Dialogue Semantics).

The focus of this lecture is on the formalisation of the basic description models. Linguistic issues are hence described by ways of phrase structure (introducing the concept of attribute-value pairs), tagset, item-and-arrangement morphology, word structure, dependency and subcategorization, etc. The theme of Dialogue Semantics is especially important, as up-to-date information systems make use of dialogue systems.

There are three written examinations during the term: Syntax A/B/C, Morphology (containing 25% questions on Syntax), and Semantics (containing each 12-20% questions on Syntax and Morphology). There is no specific book offered, instead, a manually compiled lecture script is provided in parts during the term.

Lecture (3) moves forward from general linguistic models to tools and applications where these models are implemented. Again, Syntax, Morphology, and Semantics are the main themes (feature-based grammars, finite state technology, probabilistic tagging, evaluation methodologies). Contained in this lecture are the differences between theoretical approaches and their possible implementations. The course structure is otherwise parallel to (2).

(4) In the seminars, the lecturers focus inter alia on machine translation or terminology extraction (both mainly directed at ICT students), e-Dictionaries, Word Nets, and Corpus Linguistics. Practical Courses currently concern the programming with scripting languages (perl/python) focussing on the manipulation of text data (e.g. boilerplate removal, automated metadata extraction) and on research on large amounts of text. Second, there is a regular practical course on Corpus Linguistic studies. For the latter, data from the web is utilized for the compilation of own corpora

containing texts about specific phenomena and for examining those via their linguistic features (like, for example, forum texts containing points of view on a product).

As prepared in the lectures of "Einführung in die Maschinelle Sprachverarbeitung" (3), we constantly enhance in (4) the linking of the contents of other courses of the institute with NLP issues discussed in our courses, like e-Dictionaries and Usability; Corpus Linguistics and Man-Machine Interaction; and NLP technologies, evaluation methods and Information Retrieval.

3.2 Tutorials for (2) and (3)

Weekly tutorials complement the lectures (2) and (3). For (2), groups of about 20-30 students each are to prepare their own solutions to exercises handed out during the lectures. In (3), the students get access to NLP-tools. Their training includes hands-on exercises in the use of WebLicht (`https://weblicht.sfs.uni-tuebingen.de`), and several online demos, like, for example the dependency parser Parzu (`https://pub.cl.uzh.ch/demo/parzu/`).

In all tutorial sessions, possible solutions to the exercises are presented and discussed with the students. According to the regulations of our university, we cannot force attendance to these tutorials, however, we inform our students regularly that "presence really works" (Schulmeister, 2017).

3.3 Screencasts for (2) and - in future - (3)

In our experience, a rather big issue for the students is practicing and rehearsing (computational) linguistic terminology with their lecturers, especially in the base lecture (2, *Language Description for Language Technology*). We thus decided to introduce a set of screencasts in the summer term of 2017, provided along with this lecture in order to give them "permanent access to the training material and therefore the opportunity for students constantly review the material" ((Lisetskyi:2015, , p. 32)). The screencasts[1] an they re-introduce each of the linguistic fields from a new perspective (compared to the lectures). Each one consists of a set of slides which are provided as .pdf and as a downloadable .mp4 file showing the slides plus a recorded voice describing them. Literature

mentioned in the screencast is summarised in a bibliography on the last slides.

So far, there are no screencasts for NLP issues available, but for the upcoming winter term 17/18 we plan to extend the device to accompany lecture (3): For each topic (tagging, morphological analysis, parsing), we will provide two screencasts; one (slides) describing the theoretical model of a specific implementation and a second, showing a demo, i.e. a step-by-step demonstration of the tool's usage. All tools described will be available on the internet to ensure that the students can watch the screencasts and try out the tools whenever and wherever they want. The efforts to produce a screencast are slightly higher than preparing for a lecture; making use of a commercial software allowing for easy postproduction however speeds up the process. On average,the production of such two screencasts takes one working day.

In blended learning, a key issue is to provide the transferrable knowledge from several perspectives, so that at least most of individual backgrounds and learning approaches of the students may be covered. While in the lectures the order of presentation usually leads from a term via its definition to the respective examples, the screencasts always begin with example words or sentences, followed by a categorisation and, lastly, the term(s) describing them. A constant colour scheme is used to ease the identification of each item. The sections are kept short and at the end of each one, a summary slide repeats the terms and their meaning. The slides are kept as simple as possible, so that the listener can focus on the audio part describing them; hence animation is kept to a minimum. We try to keep each screencast shorter than 15 minutes.

As the knowledge transferred in lectures, scripts, slides and screencasts is based on scientific publications, we obviously cite the background literature. We aim at our students reading this literature, too, so our goal is to rather work with a small amount of books and we choose books that are either available in the library or not too expensive to buy. A publication regularly referred to is for example (Bußmann, 2008) with the aim of directing students to use proper terminological dictionaries. Additionally, specific textbooks (like (Müller, 2009) or (Wöllstein-Leisten et al., 1997)) are cited and explained as further reading hopefully encouraging our students to read the original literature and to mention it in discussions.

---

[1] Our screencasts are produced with the commercial Software Camtasia `https://www.techsmith.com/video-editor.html`.

4 Experiences / Evaluation

Generally, the NLP teaching concept is well accepted by the students and we receive a overall positive feedback on the lectures, both face-to-face, and in the courses' evaluation which are written anonymously. About half of the participants of (2) afterwards choose lecture (3) as well, though this is not mandatory.

Seminars and Practical Courses on Corpus Linguistics are usually overbooked (around 30 bookings of IIM students for 20 available posts), we assume that this is because the theme of "Big Data" is mentioned in the description of several of the courses – this theme is of a high interest for students of Information Management.

The programming courses on scripting languages are rather small (usually 8-12 participants), as most IIM students are hesitant to learn any form of programming. The courses however regularly receive very good evaluation results and it is especially mentioned by the students that they prefer the low number of participants.

Concerning feedback on the newly introduced screencasts, so far, only three screencasts have been provided to support lecture (2), but the students frequently refer to them in the discussions and they were mentioned very positively in the course evaluation (about 50% of the students took part in the written evaluation). This response is encouraging; hence we plan to extend the number of screencasts and to add them also to the higher level courses, as described in paragraph 3.3. In order to reach a wider public, the screencasts will soon be downloadable for free from a University Hildesheim website which can be read from outside the University. This is planned because we consider especially such theme-introducing screencasts to be usable for other, similar courses, as well.

Additionally, we are currently developing a glossary of grammatical terms in a new format. The idea for this glossary stems from the evaluation of the beta version of an online project called "German as a (scientific) language" (eDaF, `https://www.uni-hildesheim.de/eDaF`). The glossary will add written definitions of terms to the screencasts. We hope that they will assist the students in finding definitions and even more examples for the terms they have to learn. This glossary will include graphical elements (like an ontology of terms, or dependency graphs), see (Faass et al., 2017).

5 Summary and Future work

Teaching NLP should not be begun before teaching linguistics, in a general introduction followed introducing the methods of formal modelling. Such modelling forms the basis for understanding NLP applications or implementations. Our teaching methods follow the principle of blended learning, as the knowledge transfer is available on different paths from which the individual student can choose. One newly introduced instrument are screencasts that are provided in addition to the live oral presentation. The screencasts present the same knowledge as was provided in the lecture, but from a different perspective, thereby referring to the same literature.

So far, we have not specifically evaluated the use of our screencasts, but the feedback we have received during the past term was very encouraging, hence we plan to extend their provision to the lecture on *Language Description for Language Technology* and to additionally produce a number of screencasts for the follow-up lecture *Introduction to Natural Language Processing* (here: application descriptions in theory and practice and demos). At the end of term 17/18, we will ask our students in a survey about their utilization of the screencasts and their point of view (whether they considered them helpful). In order to enable other lecturers (also of other Universities) to make use of them, all screencasts will be made available freely on the web from autumn 2017 on.

References

Bußmann, Hadumod. 2008. *Lexikon der Sprachwissenschaft*. 4. Auflage Stuttgart: Kröner.

Dickson, K.A.; Stephens, B.W. Standing room only: faculty intervention increases voluntary lecture attendance and performance for disadvantaged year 1 Bioscience students. *Higher Education Pedagogies* (1) 1: 1-15 DOI: 10.1080/23752696.2015.1134196

Faaß, Gertrud; Bohle, Ulrike; Quahlo, Jasmin. masked 2017. *The Grammatical Glossary: a glossary of German grammar terms.* accepted as a poster at eLex 2017 conference in Leiden, September 2017

Graefen, Gabriele; Liedke, Martin. 2012. *Germanistische Sprachwissenschaft. Deutsch als Erst-, Zweit- oder Fremdsprache.* Stuttgart: UTB.

Müller, Horst M. (ed). 2009. *Arbeitsbuch Linguistik.* 2. Auflage. Paderborn/München/Wien/Zürich: Schöning.

Lisetskyi, K.A. 2015. Blended Learning Model in the System of Higher Education *Advanced Education* 2015/4:32-35

Neumann, Petra. 2005. A closer look at blended learning-parameters for designing a blended learning environment for language teaching and learning. *ReCALL*, 17(2):163-178.

O'Connor, Christine, Mortimer, Dennis, Bond, Sue. 2011. Blended Learning: Issues, Benefits and Challenges *IJES*, 19(2).

Schulmeister, R. 2017. Presence and Self-Study in Blended Learning. *eleed*, Issue 12. `https://eleed.campussource.de/archive/12/4502` [23-08-17]

Wöllstein-Leisten, Angelika; Heilmann, Axel; Stepan, Peter; Vikner, Sten. 1997. *Deutsche Satzstruktur. Grundlagen der syntaktischen Analyse.* Tübingen: Stauffenburg Verlag Brigitte Narr GmbH.