The Lacunae of Danish Natural Language Processing

Andreas Kirkedal[†] 🇩🇰
ITU Copenhagen & Interactions LLC
Denmark
anki@itu.dk

Barbara Plank 🇩🇰
ITU Copenhagen
Denmark
bplank@itu.dk

Leon Derczynski 🇩🇰
ITU Copenhagen
Denmark
ld@itu.dk

Natalie Schluter 🇩🇰
ITU Copenhagen
Denmark
natschluter@itu.dk

Abstract

Danish is a North Germanic language spoken principally in Denmark, a country with a long tradition of technological and scientific innovation. However, the language has received relatively little attention from a technological perspective. In this paper, we review Natural Language Processing (NLP) research, digital resources and tools which have been developed for Danish. We find that availability of models and tools is limited, which calls for work that lifts Danish NLP a step closer to the privileged languages.

Dansk abstrakt: Dansk er et nordgermansk sprog, talt primært i kongeriget Danmark, et land med stærk tradition for teknologisk og videnskabelig innovation. Det danske sprog har imidlertid været genstand for relativt begrænset opmærksomhed, teknologisk set. I denne artikel gennemgår vi sprogteknologi-forskning, -ressourcer og -værktøjer udviklet for dansk. Vi konkluderer at der eksisterer et fåtal af modeller og værktøjer, hvilket indbyder til forskning som løfter dansk sprogteknologi i niveau med mere privilegerede sprog.

1 Introduction

Danish is the majority language of the Kingdom of Denmark, a country of around six million people, with five written languages across its many islands (others including Færøysk, Kalaallisut, Tunumiit oraasiat, and Borrinjholmsk (Der-czynski and Kjeldsen, 2019)). Despite its privileged place in the world, Denmark has not kept up pace with comparable countries in developing language technology. Few systems are designed explicitly for Danish; rather, general-purpose systems might be run on Danish and results produced for it as a by-product of larger studies. This supposes having adequate developed datasets. As a result, language technology does not have as prominent a place in Denmark as it might in other countries. This paper gives an overview of NLP models, tasks and datasets for Danish.

Traditionally, the country has created corpora, lexicographic resources, and other symbolic knowledge through government sponsorship. This has led to excellent research at the Dansk Sprognævn (dsn.dk) and by CLARIN DK (clarin.dk), who have both consistently produced volumes of quality Danish data within their remit. This paper examines NLP from perhaps the opposite direction: our study is task-driven instead of corpus-driven, meaning we pragmatically consider what NLP technology exists, how it is represented in the scope of Danish, and, where appropriate, what might help improve the situation.

While Denmark has multiple languages (as above), Danish also has multiple language variants – for example ømålsdansk, which encompasses all from the absence of stød (Hansen, 1943; Basbøll, 2005) in fynsk, to københavnsk, with its quirks, like the use of *forrasten* in place of standard Danish *forræsten* (Institut for Dansk Dialektforskning, 1992-). This study ignores these variants, focusing on standard Danish. We recognise that this choice perpetuates the erosion of other tongues within Denmark but, at the same time, we are aware of how the high prevalence of English in the country similarly erodes access to good NLP for Danish users – and addressing the lacunae of latter is the primary concern for this paper.

---

†: Research Scientist at Interactions LLC.
🇩🇰: These authors contributed to the paper equally.

We present an overview of the status of Danish on a sample of NLP tasks drawn from the "NLP Progress" list,[1] automatic speech recognition and speech synthesis, organized thematically. This work considers speech to be natural language and applications such as automatic speech recognition and speech synthesis as NLP tasks.

2 Syntactic Tasks

Starting at the most basic linguistic hierarchy is often to identify the syntactic structure of a sentence.

2.1 Part-of-speech tagging

PoS tagging is the task of assigning abstract basic syntactic categories to every token. PoS tagging is one of the cornerstone NLP tasks and typically one of the first to be addressed for a new language. Consequently, PoS tagging schemes and corpora have emerged for a variety of languages, including Danish (Bilgram and Keson, 1998).

Typically, each annotation effort developed their own annotation guidelines. Early work on Danish included over 100 fine-grained PoS tags (Bilgram and Keson, 1998). A recent initiative, the Universal Dependencies (UD) (Nivre et al., 2016), initiated a new broadly-adopted model to homogenize prior diverging efforts. By sacrificing detail for standardisation, UD proposes a unified annotation scheme for syntactic annotation of dependency trees including PoS tags and morphological features, which maximizes parallelism between languages while still allowing for language-specific annotations. For PoS, the UD scheme consists of 17 universal PoS tags.[2] The latest UD release (v2.4) covers 83 languages. UD has been widely adopted in both academia and industry (Nivre et al., 2016; Bohnet et al., 2018).

Two existing Danish-specific PoS taggers exist under restricted access (Asmussen, 2015). They include mostly rule-based systems accessible via an online interface: the Brill PoS tagger developed by the Centre for Language Technology[3] and a tagger developed by GrammarSoft. In contrast, many general purpose tagging tools are widely available as open-source taggers.[4] The current best systems rely on deep learning implementing bidirectional LSTM architectures (Plank et al.,

2016; Bohnet et al., 2018). They reach accuracies in the high 90s for Danish, i.e 96% on UD Danish (Plank et al., 2016). Contrary to major languages such as English, there is a lack of data for PoS annotated data for non-canonical domains like social media or specialized medical data.

2.2 Dependency Parsing

Dependency parsing is the task of identifying the syntactic structure of a sentence. In dependency parsing, the syntactic structure is expressed as a set of bilexical head-modifier relationships called *dependencies*, e.g., $subj(Anna, sings)$. The set of dependencies forms a tree structure, thereby yielding a structured prediction problem.

The first Danish treebank is the Copenhagen Dependency Treebank (CDT) (Kromann et al., 2003). It consists of 100k tokens of syntactically-annotated data from the Parole corpus (Bilgram and Keson, 1998). The Danish UD treebank (`Danish-DDT`) is a conversion of the Copenhagen Dependency Treebank to UD (Johannsen et al., 2015). In recent evaluations, labeling accuracies of 86% were reported for Danish UD dependency parsing (Zeman et al., 2018), mainly over news articles. Overall, Danish dependency parsing has received the most attention.

3 Semantic Tasks

The processing tasks that depend on the meanings of a target text are gathered in this section. While a broad area of NLP, including meaning representation, commonsense reasoning, automatic summarization, spatial and temporal information extraction, and linguistic inference, we focus areas where some work on Danish exists: recognising name mentions and supersenses, handling clinical text, and sentiment extraction.

3.1 Named Entity Recognition and Senses

Picking up on specifically named items, like names of people, places and organizations, can lead to useful analyses; this is called Named Entity Recognition (NER). For some genres and languages, NER has advanced to high accuracies (e.g. English Newswire). For others, the technology is less advanced. It is a more coarse-grained task than sense tagging which has received attention in Danish (Alonso et al., 2015; Pedersen et al., 2015).

Many NER results for Danish are outdated and based on closed, systems. E.g., Bick (2004) offers

---

[1] https://nlpprogress.com/
[2] universaldependencies.org/u/overview/morphology.html
[3] https://cst.dk/online/pos_tagger/
[4] https://github.com/bplank/bilstm-aux/

details of a system trained on 43K tokens but reports no F1. One has to pay for this tool and the data is not open. Johannessen et al. (2005) mention efforts in Danish NER but the research lies behind a paywall that the authors do not have access through, and we failed to find other artefacts of this research. More recently, Derczynski et al. (2014) describe a dataset used to train a recognizer that is openly available in GATE (Cunningham et al., 2012). Current efforts focus on addressing the problem of data sparsity and on providing accessible tools (Plank, 2019; Derczynski, 2019), including as part of the ITU Copenhagen open tool set for Danish NLP.[5]

In contrast, for English, F1 scores are in the mid-90s (e.g., 94.03 from Chiu and Nichols (2016)). Researchers have since moved on to more exotic challenges, such as nested entities, emerging entities, and clinical information extraction (Katiyar and Cardie, 2018; Derczynski et al., 2017; Wang et al., 2018).

To improve Danish NER seems simple: we need open tools and annotated data. Fortunately, the landscape for Danish NER is somewhat barren, and so first movers have an advantage. Openly contributing such a dataset to a shared resource would mean that Danish NER would be included in multilingual NER exercises, thus enabling the rest of the world to also work on improving entity recognition for Danish.

3.2 Clinical IE

The language used in biomedical and clinical applications has its own nuance. Technical terms abound, and dialects vary between specialisations and even from institution to institution. Patient record notes have the potential for particularly broad variations: they are uncurated, they are not designed for publication, and the target audience tends to be quite similar to the target author, thus permitting greater use of idiosyncratic language. These factors make text in this domain hard to deal with for standard tools. They also make it difficult to use transfer approaches from other languages. For example, while one might reasonably be able to use belles lettres in English to better process belles lettres in Danish, the idiosyncracies in clinical notes mean that one language's clinical note data is unlikely to hugely help understanding clinical notes in other languages.

Danish clinical NLP lags behind that for other languages, even when English is taken out of the picture, with for example four times as many Pubmed references to Swedish clinical processing, and twice as many to Finnish, than exist for Danish clinical NLP (Névéol et al., 2018). Work relies on older technology, not exploiting the higher performance of deep learning (Eriksson et al., 2013). Efforts to improve on this situation are hampered by the data being tightly closed to NLP researchers, compared to the situation in Sweden and Finland – this despite Denmark having an unusually rich archive of clinical data, which is "gathering dust" (Reiermann and Andersen, 2018).

There is limited Danish clinical data (Pantazos et al., 2011), but basic tasks such as entity recognition are not yet in place. Adverse drug reaction extraction tools have been built (Eriksson et al., 2013), achieving an F1 of 0.81 on psychiatric hostpial patient records, compared to an F1 of 0.87 for English on the more difficult task of multi-genre records (Huynh et al., 2016). Clinical timeline extraction (Sun et al., 2013; Bethard et al., 2016) is absent for Danish.

To improve the situation for people whose medical data is stored in Danish, both the institutional access problem and the technology development problems need to be addressed. Fortunately, research into clinical NLP for other languages is quite advanced, making it easier to catch up.

3.3 Sentiment Extraction

Sentiment analysis is a long-standing NLP task for predicting the sentiment of an utterance, in general or related to a target (Liu, 2012). It has been investigated for non-formal text, which presents its own hurdles (Balahur and Jacquet, 2015), leading to a series of shared tasks (Rosenthal et al., 2015).

The `afinn` tool (Nielsen, 2011) performs sentiment analysis using a lexicon consisting of 3552 words, labelled with a value between -5 (very negative) and 5 (very positive). This approach only considers the individual words in the input, and therefore the context is lost.

Full-text annotations for sentiment in Danish have appeared in previous multilingual work, including systems reaching F-scores of 0.924 on same-domain Trustpilot reviews and 0.462 going across domains (Elming et al., 2014). Alexandra Institute offer a model[6] based on Facebook's

---

LASER multilingual sentiment tool.[7] This is total of Danish sentiment text tools, and all are included incidentally as part of multilingual efforts.

4 Machine Translation

Machine translation (MT) is the automatic translation from one language to another. MT typically thrives on sentence-aligned data, where sentences in the source language are paired with their translation in the target language. Tools specifically designed in Denmark for Danish are not open and often only translate one way;[8] this makes them impossible to benchmark.

On the other hand, it is rare that translation tools include Danish in evaluations. Popular pairs are en-fr, en-de, en-zh and en-ja, which tend to be present in most large-scale research exercises (Johnson et al., 2017; Chen et al., 2018). When Danish does appear, it is typically in order to make a linguistic point, rather than improve MT for Danish-speakers (Vanmassenhove et al., 2018). However, even given that, there is a relatively large amount of Danish parallel text (that MT relies on): Opus[9] reports 63M sentences for English-Danish, 70M for English-Swedish, 117M for English-German, and 242M for English-French. A large amount of the Danish data comes from colloquial, crowdsourced sites like Open-Subtitles.net and Tatoeba. Just as it's incidental that Danish is included in these (i.e. their translations is not purpose-created for Danish, which is a signal of quality), there are also no dedicated Danish parallel texts listed on CLARIN.eu.[10] The result is thus that Danish MT is missing focused technology, and focused corpora, specifically designed to give correct Danish translations.

5 Speech Technology

Automatic speech recognition (ASR) converts spoken utterances to text. Converting text to spoken utterances is known as speech synthesis or text-to-speech (TTS) systems.

5.1 Automatic speech recognition

Danish ASR has received limited attention from a research perspective. In terms of data, Danish should be considered a medium-resource language largely due to the access to the open-domain speech corpus known as Språkbanken,[11] which contains 300 hours of phonetically-balanced ASR training data and 50 hours of test data – as well as data for telephony and dictation. The data is read-aloud speech which assures a good correspondence between text and speech. However, this genre does not contain examples of many issues in realistic speech like dysfluencies, restarts, repairs and foreign accents. ELRA hosts the Speech-Dat/Aurora, EUROM1 and Collins data collections behind a paywall, but these do not contain a substantial amount of spontaneous speech; access to realistic spontaneous speech is extremely limited for Danish languages. This is a barrier to research and development for Danish ASR. Creating these resources from scratch is expensive and cannot be undertaken by start-ups, SMEs or single research groups without substantial backing.

In terms of available software or systems, a speech recogniser training recipe based on the Kaldi toolkit (Povey et al., 2011) is available online.[12] This is a hybrid DNN-HMM system that requires a phonetic transcription, but if we desire to train end-2-end ASR systems, phonetic transcription is not necessary and we can take an off-the-shelf toolkit like OpenSeq2Seq (Kuchaiev et al., 2018) and train an off-line system.[13] Google, Nuance, IBM and Danish companies like MIRSK, Dictus and Corti develop Danish ASR; Dictus and Mikroværkstedet also have TTS solutions. Dictus recently released Dictus Sun[14] which will be used at the Danish parliament to draft speech transcriptions.

ASR system performance depends on language models. As speech genre is important for acoustic model performance, so language models trained on newswire, Wikipedia, Twitter data or similar will not work as well as language models trained on speech transcriptions. Dictus Sun has access to 11 years of transcribed speeches and so may work well for monologues in that domain, but we have not been able to test the system and cannot know its performance on spontaneous speech.

A lot of medium quality transcribed data is better than a little perfectly transcribed data and creating more data rather than correcting existing transcriptions provides better performance (Sperber et al., 2016; Novotney and Callison-Burch,

---

[7]See https://github.com/facebookresearch/LASER
[8]See https://visl.sdu.dk/visl/da/tools/
[9]See http://opus.nlpl.eu/
[10]See www.clarin.eu/resource-families /parallel-corpora

[11]See github.com/fnielsen/awesome-danish for links.
[12]github.com/kaldi-asr/kaldi/tree/master/egs/sprakbanken.
[13]*Offline* means it cannot recognise speech in real-time.
[14]https://www.dictus.dk

2010). This was used to create the Fisher corpus, a standard benchmark (Cieri et al., 2004). We recommend this approach, coupled with release of publicly-owned parallel data (e.g. subtitles & audio from Danmarks Radio archives; Danish parliament speeches with transcriptions).

5.2 Speech synthesis.

The synthesisers available online are eSpeak and Responsive Voice.[15] Språkbanken contains a section of data that can be used to train a speech synthesiser. Recently, toolkits to train DNN-based speech synthesisers have become available online[16] because they can be trained on aligned speech and text data like ASR systems, but we are not aware of any systems or recipes to train Danish speech synthesisers. A first step would be to develop a synthesiser on the TTS part of Språkbanken and then the ASR part.

6 Discussion and Conclusion

This paper discussed a range of NLP tasks and available technologies. It is a not an exhaustive survey of Danish NLP tools: good resources and resource lists can be found out on the web. Rather, we focus on academic research and pressing tasks.

Danish language technology remains nascent. Corpora are somewhat available, but not guided by modern technological advances. The argument of a national report on language technology, parallel to and independent of this paper, was that more data is needed (DSN, 2019). In the era of deep learning, which a major part of contemporary NLP relies upon, we need huge datasets. These do not exist on the same scale as in privileged languages. Danish language text needs to be annotated, but because in the Danish context annotation is very expensive and doesn't scale (cf. e.g. annotation for the world's second language, English), one must be careful about where effort is allocated. The exact kinds of annotation must be led by modern NLP research to have the most impact, listening to advances in the field. We recommend a top-down approach, basing choices for development on those where they are found to be lacking for a certain specific applied goals. For example, modern and colloquial parallel corpora will serve to improve the standard of machine translations that Danish speakers experience daily; sentiment

and NER datasets and benchmarks for Danish will enable the innovation and technology projects that often serve to spark local industrial interest in NLP; high-vocabulary-coverage contextual embeddings for Danish will enhance performance of contemporary machine learning approaches in both research and in innovation; including Danish in NLI datasets will drive forward progress on Danish as the NLP world works on multilingual reasoning and inference. A bottom-up approach, constructing a set of resources with the eventual goal of assembling a large, complex system, risks failing to match opportunities in Denmark and the broader NLP community. We draw an analogy between these approaches and the choice of being market-led or product-led. Product-led organisations specialise in producing one kind of product and do it very well. In contrast, market-led businesses learn their market and provide what their market wants. The bottom-up approach to structuring and funding NLP research is similar to being product-led. The resources are good, but there can be a disconnect with important parts of the community, making it a risky strategy. The present lacunae are a symptom of this strategy.

We propose that Danish language technology is steered in directions that directly support and engage with the global frontier in NLP. Danish syntactic tools, Danish semantic processing, and applied Danish NLP comprise the core pillars of such a strategy. As this paper shows, much existing Danish NLP is included incidentally as part of multilingual efforts. This means that Denmark has lost ownership and control of important parts of Danish NLP, and Danish speakers risk experiencing substandard technology as a result.

In the mean time, Danish NLP – intrinsically interdisciplinary – remains absent from local research agendae and so continues to languish; it is really this technology that we need if Danish users are to enjoy the benefits that NLP can deliver.

Acknowledgments

---

[15]See `https://responsivevoice.org/`

[16]For example `github.com/NVIDIA/tacotron2`, `github.com/r9y9/deepvoice3_pytorch`, `github.com/CSTR-Edinburgh/merlin`.

References

Héctor Martínez Alonso, Anders Johannsen, Sussi Olsen, Sanni Nimb, Nicolai Hartvig Sørensen, Anna Braasch, Anders Søgaard, and Bolette Sandford Pedersen. 2015. Supersense tagging for Danish. In *Proceedings of the Nordic Conference of Computational Linguistics (NODALIDA)*, 109, pages 21–29. Northern European Association for Language Technology.

Jørg Asmussen. 2015. Survey of POS taggers. Technical report, DK-CLARIN WP 2.1 Technical Report, DSL.

Alexandra Balahur and Guillaume Jacquet. 2015. Sentiment analysis meets social media–Challenges and solutions of the field in view of the current information sharing context. *Information Processing & Management*, 51:428–432.

Hans Basbøll. 2005. *The phonology of Danish*. Oxford University Press.

Steven Bethard, Guergana Savova, Wei-Te Chen, Leon Derczynski, James Pustejovsky, and Marc Verhagen. 2016. SemEval-2016 task 12: Clinical TempEval. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 1052–1062.

Eckhard Bick. 2004. A named entity recognizer for Danish. In *Proc. LREC*. European Language Resources Association.

Thomas Bilgram and Britt Keson. 1998. The construction of a tagged Danish corpus. In *Proceedings of the Nordic Conference on Computational Linguistics (NODALIDA)*, pages 129–139. Northern European Association for Language Technology.

Bernd Bohnet, Ryan McDonald, Goncalo Simoes, Daniel Andor, Emily Pitler, and Joshua Maynez. 2018. Morphosyntactic tagging with a meta-BiLSTM model over context sensitive token encodings. In *Proc. EMNLP*. Association for Computational Linguistics.

Mia Xu Chen, Orhan Firat, Ankur Bapna, Melvin Johnson, Wolfgang Macherey, George Foster, Llion Jones, Niki Parmar, Mike Schuster, Zhifeng Chen, et al. 2018. The best of both worlds: Combining recent advances in neural machine translation. *arXiv preprint arXiv:1804.09849*.

Jason PC Chiu and Eric Nichols. 2016. Named entity recognition with bidirectional LSTM-CNNs. *Transactions of the Association for Computational Linguistics*, 4:357–370.

Christopher Cieri, David Miller, and Kevin Walker. 2004. The Fisher Corpus: a resource for the next generations of speech-to-text. In *Proc. LREC*, pages 69–71. European Language Resources Association.

Hamish Cunningham, Diana Maynard, Kalina Bontcheva, Valentin Tablan, Cristian Ursu, Marin Dimitrov, Mike Dowman, Niraj Aswani, Ian Roberts, Yaoyong Li, et al. 2012. *Developing Language Processing Components with GATE Version 8.0: A User Guide*. University of Sheffield.

Leon Derczynski. 2019. Simple Natural Language Processing Tools for Danish. *arXiv*, abs/1906.11608.

Leon Derczynski and Alex Speed Kjeldsen. 2019. Bornholmsk Natural Language Processing: Resources and Tools. In *Proceedings of the Nordic Conference on Computational Linguistics (NODALIDA)*. Northern European Association for Language Technology.

Leon Derczynski, Eric Nichols, Marieke van Erp, and Nut Limsopatham. 2017. Results of the WNUT2017 shared task on novel and emerging entity recognition. In *Proceedings of the 3rd Workshop on Noisy User-generated Text*, pages 140–147.

Leon Derczynski, C. Vilhelmsen, and Kenneth S Bøgh. 2014. DKIE: Open source information extraction for Danish. In *Proceedings of the Demonstrations at the Conference of the European Chapter of the Association for Computational Linguistics*, pages 61–64.

DSN. 2019. Dansk sprogteknologi i verdensklasse. Technical report, Dansk Sprognævn.

Jakob Elming, Barbara Plank, and Dirk Hovy. 2014. Robust cross-domain sentiment analysis for low-resource languages. In *Proceedings of the 5th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 2–7.

Robert Eriksson, Peter Bjødstrup Jensen, Sune Frankild, Lars Juhl Jensen, and Søren Brunak. 2013. Dictionary construction and identification of possible adverse drug events in Danish clinical narrative text. *Journal of the American Medical Informatics Association*, 20(5):947–953.

Aage Hansen. 1943. Stødet i dansk. *Det Kongelige Danske Videnskabernes Selskab. Historisk-Filologiske Meddelelser*, 29.

Trung Huynh, Yulan He, Alistair Willis, and Stefan Rüger. 2016. Adverse drug reaction classification with deep neural networks. In *Proc. COLING*.

Institut for Dansk Dialektforskning. 1992-. *Ømålsordbogen*. C. A. Reitzels Forlag.

Janne Bondi Johannessen, Kristin Hagen, Åsne Haaland, Andra Björk Jónsdottir, Anders Nøklestad, Dimitris Kokkinakis, Paul Meurer, Eckhard Bick, and Dorte Haltrup. 2005. Named entity recognition for the mainland Scandinavian languages. *Literary and Linguistic Computing*, 20(1):91–102.

Anders Johannsen, Héctor Martínez Alonso, and Barbara Plank. 2015. Universal dependencies for Danish. In *Proc. International Workshop on Treebanks and Linguistic Theories (TLT)*, page 157.

Melvin Johnson, Mike Schuster, Quoc V Le, Maxim Krikun, Yonghui Wu, Zhifeng Chen, Nikhil Thorat, Fernanda Viégas, Martin Wattenberg, Greg Corrado, et al. 2017. Google's multilingual neural machine translation system: Enabling zero-shot translation. *Transactions of the Association for Computational Linguistics*, 5:339–351.

Arzoo Katiyar and Claire Cardie. 2018. Nested named entity recognition revisited. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 861–871.

Matthias T Kromann, Line Mikkelsen, and Stine Kern Lynge. 2003. Danish dependency treebank. In *Proc. International Workshop on Treebanks and Linguistic Theories (TLT)*, pages 217–220.

Oleksii Kuchaiev, Boris Ginsburg, Igor Gitman, Vitaly Lavrukhin, Jason Li, Huyen Nguyen, Carl Case, and Paulius Micikevicius. 2018. Mixed-Precision Training for NLP and Speech Recognition with OpenSeq2Seq. *arXiv preprint arXiv:1805.10387*.

Bing Liu. 2012. Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies*, 5(1):1–167.

Aurélie Névéol, Hercules Dalianis, Sumithra Velupillai, Guergana Savova, and Pierre Zweigenbaum. 2018. Clinical natural language processing in languages other than English: Opportunities and challenges. *Journal of Biomedical Semantics*, 9(1):12.

Finn Årup Nielsen. 2011. A new ANEW: Evaluation of a word list for sentiment analysis in microblogs. In *Proceedings of the ESWC2011 Workshop on 'Making Sense of Microposts': Big things come in small packages*, volume 718, pages 93–98. CEUR-WS.

Joakim Nivre, Marie-Catherine De Marneffe, Filip Ginter, Yoav Goldberg, Jan Hajic, Christopher D Manning, Ryan McDonald, Slav Petrov, Sampo Pyysalo, Natalia Silveira, et al. 2016. Universal dependencies v1: A multilingual treebank collection. In *Proc. LREC*, pages 1659–1666. European Language Resources Association.

Scott Novotney and Chris Callison-Burch. 2010. Cheap, fast and good enough: Automatic speech recognition with non-expert transcription. In *TheAnnual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 207–215. Association for Computational Linguistics.

Kostas Pantazos, Søren Lauesen, and Søren Lippert. 2011. De-identifying an EHR database-anonymity, correctness and readability of the medical record. In *MIE*, pages 862–866.

Bolette Sandford Pedersen, Sanni Nimb, and Sussi Olsen. 2015. Eksperimenter med et skalérbart betydningsinventar til semantisk opmærkning af dansk. In *Rette Ord*, pages 247–261. Dansk Sprognævns skrifter.

Barbara Plank. 2019. Cross-Lingual Transfer and Very Little Labeled Data for Named Entity Recognition in Danish. In *Proceedings of the Nordic Conference on Computational Linguistics (NODALIDA)*. Northern European Association for Language Technology.

Barbara Plank, Anders Søgaard, and Yoav Goldberg. 2016. Multilingual part-of-speech tagging with bidirectional long short-term memory models and auxiliary loss. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 412–418. Association for Computational Linguistics.

Daniel Povey, Arnab Ghoshal, Gilles Boulianne, Lukas Burget, Ondrej Glembek, Nagendra Goel, Mirko Hannemann, Petr Motlicek, Yanmin Qian, Petr Schwarz, et al. 2011. The Kaldi speech recognition toolkit. Technical report, IEEE Signal Processing Society.

Jens Reiermann and Torben K. Andersen. 2018. Guldgrube af sundhedsdata samler støv. *Mandag Morgen*.

Sara Rosenthal, Preslav Nakov, Svetlana Kiritchenko, Saif Mohammad, Alan Ritter, and Veselin Stoyanov. 2015. Semeval-2015 task 10: Sentiment analysis in twitter. In *Proceedings of the 9th international workshop on semantic evaluation (SemEval 2015)*, pages 451–463.

Matthias Sperber, Graham Neubig, Satoshi Nakamura, and Alex Waibel. 2016. Optimizing computer-assisted transcription quality with iterative user interfaces. In *Proc. LREC*, pages 1986–1992. European Language Resources Association.

Weiyi Sun, Anna Rumshisky, and Ozlem Uzuner. 2013. Evaluating temporal relations in clinical text: 2012 i2b2 Challenge. *Journal of the American Medical Informatics Association*, 20(5):806–813.

Eva Vanmassenhove, Christian Hardmeier, and Andy Way. 2018. Getting gender right in neural machine translation. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 3003–3008.

Xu Wang, Chen Yang, and Renchu Guan. 2018. A comparative study for biomedical named entity recognition. *International Journal of Machine Learning and Cybernetics*, 9(3):373–382.

Daniel Zeman, Jan Hajič, Martin Popel, Martin Potthast, Milan Straka, Filip Ginter, Joakim Nivre, and Slav Petrov. 2018. Conll 2018 shared task: Multilingual parsing from raw text to universal dependencies. In *Proceedings of the CoNLL 2018 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies*, pages 1–21.