

Sprachvarietäten- abhängige Terminologie in der neuronalen maschinellen Übersetzung: Eine Analyse in der Sprachrichtung Englisch-Deutsch mit Schwerpunkt auf der österreichischen Varietät der deutschen Sprache

Heinisch, Barbara

barbara.heinisch@univie.ac.at
Universität Wien, Österreich

Einführung

Maschinelle Übersetzung hat in den vergangenen Jahren große Fortschritte gemacht. Neuronale maschinelle Übersetzung als der aktuelle Ansatz erzielt natürlichsprachliche Ergebnisse (Hassan et al. 2018; Castilho et al. 2017; Junczys-Dowmunt et al. 2016). Allerdings können diese flüssig lesbaren Übersetzungen über inhaltliche Fehlübersetzungen hinwegtäuschen (Koehn/Knowles 2017; Forcada 2017).

Für das Training eines neuronalen maschinellen Übersetzungssystems sind große Mengen an (qualitativ hochwertigen) bilingualen Sprachressourcen erforderlich. Die Sammlung und Aufbereitung dieser Sprachressourcen, wie beispielsweise Parallelkorpora, legt damit die Grundlage für die Qualität der maschinellen Übersetzung. Allerdings gibt es auch Ansätze, die keine großen Mengen an Trainingsdaten benötigen, wie z.B. die unüberwachte maschinelle Übersetzung (Artetxe et al. 2019).

Da generische maschinelle Übersetzungssysteme mit Korpora aus unterschiedlichsten Domänen trainiert werden und daher in bestimmten Fachgebieten oft nicht die gewünschte Qualität liefern (können), werden domänenadaptierte Systeme entwickelt (Chu et al. 2017; Servan et al. 2016). Neben weiteren Besonderheiten der Fachsprache (Stolze 2009), hat insbesondere die Verwendung der korrekten Terminologie einen entscheidenden Einfluss auf die Qualitätsevaluation einer (maschinellen) Übersetzung. Im Vergleich zu regelbasierten maschinellen Übersetzungssystemen (Forcada et al. 2011; Scansani et al. 2017; Simard et al. 2007), die das Einbinden von Wörterbüchern erlauben, sind die mit dem Training des künstlichen neuronalen

Netzwerkes bei der neuronalen maschinellen Übersetzung verbundenen „Lernprozesse“ intransparenter. Daher ist auch die Einbindung von terminologischen Ressourcen schwieriger.

Terminologie als Herausforderung in der maschinellen Übersetzung (Wloka et al. 2013; Reynolds 2015) hat zu Lösungsvorschlägen geführt, die unter anderem auf der Ebene der neuronalen Dekodierung (Hasler et al. 2018) ansetzen oder die Form von Terminologieangaben als Annotationen im Ausgangstext annehmen (Dinu et al. 2019). Auch eine korpus- und fallbasierte Adaption eines generischen maschinellen Übersetzungssystems (Farajian et al. 2018; Servan et al. 2016) bietet die Möglichkeit, die zielsprachliche Terminologie in der maschinellen Übersetzung an eine Domäne anzupassen. Außerdem wird an der Anpassung der maschinellen Übersetzung an das individuelle Vokabular einer Person geforscht (Michel/Neubig 2018).

Neben der Anpassung eines maschinellen Übersetzungssystems an ein bestimmtes Fachgebiet (Chu et al. 2017), spielt auch die Berücksichtigung der Sprachvarietät eine entscheidende Rolle (Costa-jussà et al. 2018; Lakew et al. 2018). Denn Benennungen, die Homonyme sind, können einen unterschiedlichen Begriffsinhalt haben. Dies kann zu Missverständnissen, vor allem in der fachsprachlichen Kommunikation, zwischen SprecherInnen unterschiedlicher Varietäten einer Sprache führen. Im Bereich der maschinellen Übersetzung sind auch Dialekte, als nicht standardsprachliche Varietäten, ein Forschungsgegenstand, z.B. Dialekte im Schweizerdeutschen, die aufgrund der Vorherrschaft der gesprochenen Sprache über keine standardisierte Schreibweise verfügen (Honnet et al. 2018).

Demnach gilt es neben der Terminologie auch die Sprachvarietät beim Training eines (neuronalen) maschinellen Übersetzungssystems zu berücksichtigen. Obwohl sich die schriftlichen standardsprachlichen Varietäten der deutschen Sprache auch in vielen Aspekten unterscheiden (Ammon 1995), so liegt der Fokus der vorliegenden Untersuchung auf dem Bereich der Lexik, insbesondere der Terminologie.

Am Schnittpunkt zwischen Translationswissenschaft, Terminologiewissenschaft und Varietätenlinguistik kann die vorliegende Untersuchung zu den (digitalen) Geisteswissenschaften Erkenntnisse zu der Verwendung von Korpora in der maschinellen Übersetzung unter dem Aspekt der sprachvarietätenabhängigen Terminologie beitragen.

Forschungsdesign

Ausgehend von der Hypothese, dass (generische) maschinelle Übersetzungssysteme überwiegend die deutsche Standardvarietät der deutschen Sprache als Übersetzungsergebnis für Terminologie ausgeben, wurde eine Analyse in der Sprachrichtung Englisch-Deutsch mit drei (frei) verfügbaren neuronalen

maschinellen Übersetzungssystemen durchgeführt. Bei diesen maschinellen Übersetzungssystemen handelt es sich um Google Translate, DeepL sowie eTranslation. Die Wahl fiel auf diese drei Systeme, da die beiden erst genannten von professionellen ÜbersetzerInnen und Studierenden der Translationswissenschaft in Österreich verwendet werden (Heinisch/Lušický 2019; Lušický/Heinisch (unveröffentlichtes Manuskript)). Das maschinelle Übersetzungssystem eTranslation wiederum wurde von der Europäischen Kommission entwickelt und steht MitarbeiterInnen der öffentlichen Verwaltung in allen EU-Mitgliedsstaaten kostenlos zur Verfügung, wobei es einen Schwerpunkt auf EU- und EU-relevanten Texten hat (Lösch et al. 2018).

Zum Testen der Hypothese wurden drei Domänen ausgewählt, in denen sich die Standardvarietäten des Deutschen unterscheiden: Kulinarik (Schmidlin 2011), Verwaltungssprache und Rechtssprache (Wissik 2013; Lohaus 2000), wobei hier der Universitätsterminologie besondere Berücksichtigung zukommt (Heinisch-Obermoser 2014). Der Schwerpunkt wurde außerdem auf die österreichische Standardvarietät der deutschen Sprache (im Vergleich zur deutschen) in der Übersetzung aus dem Englischen gelegt. Dafür wurden einschlägige terminologische Ressourcen, die „österreichisches Deutsch“ als Ausgangssprache und – sofern vorhanden – „Englisch“ (unabhängig von der Varietät) als Zielsprache hatten, verwendet. In der Domäne der Kulinarik wurde die Liste der österreichischen Ausdrücke, die im Protokoll Nr. 10 über die Verwendung spezifisch österreichischer Ausdrücke der deutschen Sprache im Rahmen der Europäischen Union im Zuge des EU-Beitritts Österreichs ausverhandelt wurde, verwendet (Markhardt 2002; EU 1995). In der Domäne der österreichischen Verwaltungssprache wurde das „Fachglossar österreichische Verwaltung. Deutsch -Englisch“ dem Sprachressourcenportal Österreichs entnommen (Heinisch 2018). Die Domäne Universitätsterminologie, die sich an der Schnittstelle zwischen Rechts-, Verwaltungs- und disziplinärer Wissenschaftssprache(n) befindet, konnte mit einem Auszug aus der Terminologiedatenbank der Universität Wien (UniVieTerm) (Heinisch-Obermoser 2014, 2016) abgedeckt werden. Da es sich um originäre Ressourcen handelt, weisen diese eine unterschiedliche Anzahl an Einträgen auf: Kulinarik (23 Vorzugsbenennungen), Verwaltung (695 Vorzugsbenennungen) und Universität (779 Vorzugsbenennungen). Daher gilt es bei der Interpretation der Ergebnisse der vorliegenden Studie zu beachten, dass die Domäne der Kulinarik im Vergleich zu den beiden anderen terminologischen Ressourcen deutlich weniger Benennungen umfasst.

Die Ausgangstexte für die maschinelle Übersetzung waren daher die englischsprachigen Vorzugsbenennungen sowie sämtliche weitere zulässigen Benennungen, Abkürzungen waren ausgenommen. Die Benennungen wurden in die genannten maschinellen Übersetzungssysteme eingegeben, wobei als

Ausgangssprache Englisch und als Zielsprache Deutsch gewählt wurde.

Die Analyse der maschinellen Übersetzung erfolgte mittels der in der Literatur zur maschinellen Übersetzung weit verbreiteten Multidimensional Quality Metrics (MQM) (Lommel et al. 2014), die die Bestimmung von Fehlern und damit der Qualität einer (maschinellen) Übersetzung erlaubt. In der vorliegenden Studie kommt der Fehlerkategorie „Terminologie“ – und hier wiederum der Unterkategorie „Inkonsistenz mit Terminologiedatenbank“ – besondere Bedeutung zu.

Die Ergebnisse der maschinellen Übersetzung wurden mit der in der terminologischen Ressource genannten deutschen Vorzugsbenennung (in der österreichischen Standardvarietät) verglichen. Hierbei wurden drei Arten der Übereinstimmung unterschieden: Vollständige Übereinstimmung bedeutet, dass nur die Vorzugsbenennung (sprich die österreichische Varietät) in der Übersetzung aufscheint. Partielle Übereinstimmung bedeutet, dass bei der Angabe von zwei (oder mehr) möglichen Benennungen im Englischen (in der terminologischen Ressource angeführte Vorzugsbenennung und weitere zulässige Benennungen im Englischen) zumindest eine der Benennungen im Deutschen die österreichische Standardvarietät (Vorzugsbenennung) ist bzw. dass es geringfügige Abweichungen von der deutschen Vorzugsbenennung gibt, z.B. Groß- und Kleinschreibung, Verwendung von Bindestrichen, Ergänzungen bzw. andere Wortstellung bei Phraseologie bzw. Komposita, die dem Begriffsinhalt entsprechen usw. Keine Übereinstimmung bedeutet, dass es weder eine vollständige noch partielle Übereinstimmung gab. Dies ist der Fall, wenn entweder ausschließlich die deutsche standardsprachliche Varietät der Terminologie (und nicht die österreichische) verwendet wurde oder gänzlich andere Benennungen in der Übersetzung aufscheinen. Dies bedeutet, dass die Benennungen, die primär der deutschen Standardvarietät zugeordnet werden können, in den beiden Kategorien „partielle Übereinstimmung“ und „keine Übereinstimmung“ zu finden sind.

Ergebnisse

Insgesamt wurden 1497 Benennungen aus den drei genannten Domänen (Kulinarik, Verwaltung, Universität) in Form der englischen Vorzugsbenennungen (und der zulässigen Benennungen) mit den drei erwähnten maschinellen Übersetzungssystemen (GoogleTranslate, DeepL, eTranslation) ins Deutsche übersetzt. Danach wurde die maschinelle Übersetzung der Terminologie mit den deutschen Vorzugsbenennungen in den genannten terminologischen Ressourcen, die die sprachvarietätenabhängige österreichische Terminologie abbilden, verglichen. Dieser Vergleich in Abbildung 1, in der ausschließlich vollständige und keine partiellen Übereinstimmungen mit den Vorzugsbenennungen im

Deutschen berücksichtigt wurden, zeigt das Ausmaß der Berücksichtigung der österreichischen Standardvarietät der deutschen Sprache (in Form der deutschen Vorzugsbenennungen in den genannten terminologischen Ressourcen).

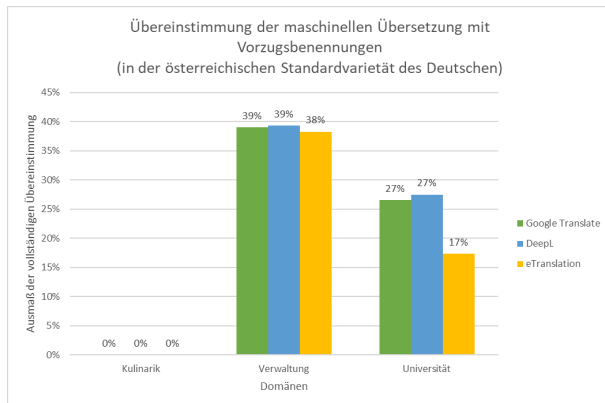


Abbildung 1: Vollständige Übereinstimmung der deutschen Übersetzung der drei verwendeten maschinellen Übersetzungssysteme mit den deutschen Vorzugsbenennungen (in der österreichischen Standardvarietät des Deutschen) in den drei terminologischen Ressourcen in den Domänen Kulinarik, Verwaltung und Universität (Prozentangaben sind gerundet)

Die Ergebnisse fallen abhängig von der Domäne unterschiedlich aus. Besonders deutlich wird dies im Bereich der Kulinarik, da hier kein einziges der untersuchten maschinellen Übersetzungssysteme die Vorzugsbenennung (im österreichischen Standarddeutsch) ausgibt. Die Analyse der Kategorien „partielle Übereinstimmung“ und „keine Übereinstimmung“ in der Kulinarik-Domäne zeigt, dass die deutsche Varietät in beinahe jeder Übersetzung der untersuchten maschinellen Übersetzungssysteme vorherrschend ist. Die einzigen Ausnahmen für die Benennungen Kren und Topfen fanden sich lediglich bei eTranslation. (Da diese jedoch nur durch die Verwendung von weiteren zulässigen Benennungen im Englischen (Synonyme) und in Verbindung mit der deutschen standardsprachlichen Terminologie als Synonyme aufgelistet wurden, gelten diese als partielle Übereinstimmung und scheinen nicht in Abbildung 1 auf.) Allerdings gilt es anzumerken, dass beispielsweise DeepL alternative Übersetzungen zu einer Benennung mittels Dropdown-Menü anbietet. In dieser Liste wurde in vielen Fällen die österreichische Varietät der Benennung als ein Listenelement – wenn auch teils weit unten – angezeigt. Das Ergebnis von eTranslation war überraschend, da im Protokoll 10 des EU-Beitrittsvertrags Österreichs die terminologischen Besonderheiten im Bereich der Kulinarik spezifiziert wurden und von einer umfassenderen Verwendung von österreichischen Ausdrücken hätte ausgegangen werden können.

In der Domäne der Verwaltungssprache unterschieden sich die drei untersuchten Systeme kaum (je ca. 40% vollständige Übereinstimmung). Es wurde ein Überhang der deutschen Varietät über alle drei maschinellen Übersetzungssysteme hinweg bemerkt. Nur bei wenigen Benennungen gaben alle drei Systeme zugleich die deutsche Varietät bzw. eine andere Übersetzung aus. Dies betraf in erster Linie Benennungen, die Bezirk, Gemeinde oder Magistrat enthielten. Bei anderen Benennungen waren die Ergebnisse variabel.

In der Domäne der Universitätsterminologie schließlich wurde erstmals auch ein Unterschied zwischen den maschinellen Übersetzungssystemen deutlich. GoogleTranslate und DeepL mit je 27% vollständiger Übereinstimmung gaben deutlich mehr Vorzugsbenennungen (in der österreichischen Standardvarietät) aus als eTranslation (mit 17%). Des Weiteren war grundsätzlich ein deutlicher Überhang der deutschen Standardvarietät bei den untersuchten Benennungen zu beobachten. Außerdem gab es in dieser Domäne auch die meisten Ambiguitäten in den Übersetzungen, da es unter anderem deutschsprachige Benennungen (oder Paraphrasierungen) gab, die nicht der Universitätsterminologie zugeordnet werden können und daher „keine Übereinstimmungen“ darstellen. Als Beispiel hierfür kann das *Sammelzeugnis* (*transcript of records*) genannt werden, das mit *Abschrift von* bzw. *der Aufzeichnungen* oder *Transkript der Aufzeichnungen* übersetzt wurde. Interessanterweise übersetzten zwei der drei Systeme die auch im Englischen auf Deutsch belassene Funktion *Studienpräses* (entweder mit *Studienpräsenzen* oder mit *Studienbeihilfen*). Daher ist davon auszugehen, dass vor allem in dieser Domäne Missverständnisse durch die maschinelle Übersetzung entstehen können. Das schlechte Ergebnis in der Domäne der Universitätsterminologie kann allerdings auch auf die verwendete Ressource und darauf zurückzuführen sein, dass die Universität Wien teilweise ihre eigene deutschsprachige Terminologie (*Corporate Terminology*) prägt.

Zusammenfassend kann gesagt werden, dass es bei den drei (frei) verfügbaren neuronalen maschinellen Übersetzungssystemen einen signifikanten Überhang an standardvarietätenabhängiger Terminologie gibt, die der deutschen Varietät des Deutschen zugeordnet werden kann.

Diskussion

Durch die Auswahl der Daten beim Zusammenstellen der Korpora für das Training eines (neuronalen) maschinellen Übersetzungssystems kommt es bereits zu einer Verzerrung in eine bestimmte Richtung. Diese Verzerrung kann gewollt erfolgen, z.B., wenn ein maschinelles Übersetzungssystem an eine bestimmte Domäne oder bestimmte Bedürfnisse angepasst werden soll oder sie kann ungewollt aufgrund der begrenzten Verfügbarkeit von bzw. des Zugangs zu Sprachressourcen

geschehen. Außerdem sind manche Varietäten in Sprachressourcen aufgrund der geringeren Anzahl der SprecherInnen unterrepräsentiert. Somit kann die Auswahl (und Verfügbarkeit) von Sprachressourcen Auswirkungen auf die sprachliche Vielfalt der maschinellen Übersetzung haben.

Für die digitalen Geisteswissenschaften sind diese Ergebnisse insofern relevant, da sie eine Reflexion über die Ergebnisse von maschineller Übersetzung erlauben, sowie ein kritisches Hinterfragen, Interpretieren und Verwenden von maschinellen Übersetzungssystemen ermöglichen.

Weiterführende Studien

Weiterführende Studien, die sich mit Sprachvarietäten in der maschinellen Übersetzung beschäftigen, sollten neben der Terminologie auch andere Unterscheidungsmerkmale in der plurizentrischen deutschen Standardsprache sowie Non-Standard (Neubarth/Trost 2017) berücksichtigen. Außerdem könnte die Übersetzung von kohärenten (originären) Texten, die relevante Terminologie enthalten, mit maschinellen Übersetzungssystemen ein anderes Ergebnis liefern. Darüber hinaus könnten sich weiterführende Studien neben der MQM-Fehlerkategorie Terminologie der Untersuchung der übrigen MQM-Kategorien, wie Genauigkeit und Stil widmen, da diesen neben der Terminologie in der Fachsprache ebenfalls Bedeutung zukommt. Des Weiteren könnte ein eigens mit der österreichischen Varietät des Deutschen trainiertes maschinelles Übersetzungssystem ebenfalls für den Vergleich genutzt und die Ergebnisse mit einem Sprachmodell für die österreichische Varietät des Deutschen erneut analysiert werden.

Schlussfolgerung

Durch die Auswahl der Daten und die Menge der zur Verfügung stehenden Sprachressourcen für das Training von maschinellen Übersetzungssystemen kann es zu Verzerrungen in maschinellen Übersetzungen kommen, die die sprachlichen Spielräume und die sprachliche Vielfalt einschränken. (Standard-)Varietäten einer Sprache stellen aktuell eine Herausforderung im Bereich der neuronalen maschinellen Übersetzung dar. Dies konnte in der vorliegenden Studie im Bereich der Terminologie für die österreichische Standardvarietät des Deutschen in den Gebieten Kulinarik, Verwaltungssprache und Universitätsterminologie bestätigt werden, da die untersuchten (generischen) maschinellen Übersetzungssysteme eine Bevorzugung der deutschen Standardvarietät zeigten. Demnach gilt es, nicht nur die Sprachvarietät, sondern auch die sprachvarietätenabhängige Terminologie beim Training maschineller Übersetzungssysteme zu berücksichtigen.

Bibliographie

Ammon, Ulrich (1995): *Die deutsche Sprache in Deutschland, Österreich und der Schweiz: Das Problem der nationalen Varietäten*. Berlin / New York: De Gruyter.

Artetxe, Mikel / Labaka, Gorka / Agirre, Eneko (2019): „An Effective Approach to Unsupervised Machine Translation“, in: *Proceedings of the 57 th Annual Meeting of the Association for Computational Linguistics. Florence, Italy, July 28 - August 2, 2019*, Association for Computational Linguistics 194–203.

Castilho, Sheila / Moorkens, Joss / Gaspari, Federico / Calixto, Iacer / Tinsley, John / Way, Andy (2017): „Is Neural Machine Translation the New State of the Art?“, in: *The Prague Bulletin of Mathematical Linguistics* 108: 109-120.

Chu, Chenhui / Dabre, Raj / Kurohashi, Sadao (2017): „An Empirical Comparison of Domain Adaptation Methods for Neural Machine Translation“, in: Barzilay, Regina / Kan, Min-Yen (Hrsg.): *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Short Papers)*. Stroudsburg, PA, USA: Association for Computational Linguistics 385–391.

Costa-jussà, Marta R. / Zampieri, Marcos / Pal, Santanu (2018): „A Neural Approach to Language Variety Translation“, in: *Proceedings of the Fifth Workshop on NLP for Similar Languages, Varieties and Dialects (VarDial 2018)* 275–282.

Dinu, Georgiana / Mathur, Prashant / Federico, Marcello / Al-Onaizan, Yaser (2019): „Training Neural Machine Translation To Apply Terminology Constraints“, in: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* 3063–3068.

EU (1995): *Beachtung österreichischer Ausdrücke. Protokoll Nr. 10 über die Verwendung spezifisch österreichischer Ausdrücke der deutschen Sprache im Rahmen der Europäischen Union* https://ec.europa.eu/info/sites/info/files/austrian_expressions_de.pdf [letzter Zugriff 11. November 2019].

Farajian, M. Amin / Bertoldi, Nicola / Negri, Matteo / Turchi, Marco / Federico, Marcello (2018): „Evaluation of Terminology Translation in Instance-Based Neural MT Adaptation“, in: Pérez-Ortiz, Juan Antonio / Sánchez-Martínez, Felipe / Esplá-Gomis, Miquel / Popovič, Maja / Rico, Celia / Martins, André / van den Bogaert, Joachim / Forcada, Mikel L. (Hrsg.): *Proceedings of the 21 st Annual Conference of the European Association for Machine Translation. 28-30 May 2018, Universitat d'Alacant, Alacant, Spain* 149–158.

Forcada, Mikel L. (2017): „Making sense of neural machine translation“, in: *Translation Spaces* 6:2 291–309.

Forcada, Mikel L. / Ginestí-Rosell, Mireia / Nordfalk, Jacob / O'Regan, Jim / Ortiz-Rojas, Sergio / Pérez-Ortiz, Juan Antonio / Sánchez-Martínez, Felipe / Ramírez-Sánchez, Gema / Tyers, Francis M. (2011): „Apertium: a free/open-source platform for rule-based

machine translation“, in: *Machine Translation* 25:2 127–144.

Hasler, Eva / Gispert, Adrià De / Iglesias, Gonzalo / Byrne, Bill (2018): „Neural Machine Translation Decoding with Terminology Constraints“, in: *Proceedings of NAACL-HLT 2018, New Orleans, Louisiana, June 1 - 6, 2018* 506–512.

Hassan, Hany / Aue, Anthony / Chen, Chang / Chowdhary, Vishal / Clark, Jonathan / Federmann, Christian / Huang, Xuedong / Junczys-Dowmunt, Marcin / Lewis, William / Li, Mu / Liu, Shujie / Liu, Tie-Yan / Luo, Renqian / Menezes, Arul / Qin, Tao / Seide, Frank / Tan, Xu / Tian, Fei / Wu, Lijun / Wu, Shuangzhi / Xia, Yingce / Zhang, Dongdong / Zhang, Zhirui / Zhou, Ming (2018): *Achieving Human Parity on Automatic Chinese to English News Translation* <https://arxiv.org/pdf/1803.05567>

Heinisch, Barbara (2018): *Dissemination of administrative terminology on Austria's language resource portal as a means of quality assurance*, Donostia / San Sebastián, Spanien: Poster presented at EAFT Terminology Summit 2018: 3M4Q: Making, Managing, Measuring Terminology. In the pursuit of Quality - Donostia / San Sebastián, Spanien, 22-23 Nov. 2018.

Heinisch, Barbara / Lušicky, Vesna (2019): „User expectations towards machine translation: A case study“, in: *Proceedings of Machine Translation Summit XVII Volume 2: Translator, Project and User Tracks*, 19–23 August, 2019, Dublin, Ireland 42–48.

Heinisch-Obermoser, Barbara (2014): „University terminology: Why it is not just higher education terminology“, in: Budin, Gerhard / Lušicky, Vesna (Hrsg.): *Languages for special purposes in a multilingual, transcultural world. Proceedings of the 19 th European Symposium on Languages for Special Purposes, LSP, 2013 8-10 July, 2013 Vienna, Austria*. Vienna: Centre for Translation Studies 429–433.

Heinisch-Obermoser, Barbara (2016): „Terminology workflows at an Austrian university aimed at collaboration, terminology awareness and joint responsibility for university terminology – a study on the University of Vienna's terminological database UniVieTerm“, in: European Association for Terminology (Hrsg.): *VIII EAFT Terminology Summit (2016), 14 Nov 2016 - 15 Nov 2016, Luxembourg. Revisions and Visions* 8.

Junczys-Dowmunt, Marcin / Dwojak, Tomasz / Hoang, Hieu (2016): *Is Neural Machine Translation Ready for Deployment? A Case Study on 30 Translation Directions*. <https://arxiv.org/pdf/1610.01108>.

Koehn, Philipp / Knowles, Rebecca (2017): „Six Challenges for Neural Machine Translation“, in: *Workshop on Neural Machine Translation, Vancouver, BC*. arXiv: 1706.03872.

Lakew, Surafel M. / Erofeeva, Aliia / Federico, Marcello (2018): „Neural Machine Translation into Language Varieties“, in: *Proceedings of the Third Conference on Machine Translation (WMT). Volume*

1: Research Papers. Brussels, Belgium, October 31 - November 1, 2018 156–164.

Lohaus, Marianne (2000): *Recht und Sprache in Österreich und Deutschland. Gemeinsamkeiten und Verschiedenheiten als Folge geschichtlicher Entwicklungen; Untersuchung zur juristischen Fachterminologie in Österreich und Deutschland*. Giessen: Fachverl. Köhler.

Lommel, Arle / Uszkoreit, Hans / Burchardt, Aljoscha (2014): „Multidimensional quality metrics (MQM): A framework for declaring and describing translation quality metrics“, in: *Revista tradumàtica: traducció i tecnologies de la informació i la comunicació* 12: 455–463.

Lösch, Andrea / Mapelli, Valérie / Piperidis, Stelios / Vasi#jevs, Andrejs / Smal, Lilli / Declerck, Thierry / Schnur, Eileen / Choukri, Khalid / van Genabith, Josef (2018): „European Language Resource Coordination: Collecting Language Resources for Public Sector Multilingual Information Management“, in: Calzolari / Nicoletta (Hrsg.): *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)* 1339–1343.

Lušicky, Vesna / Heinisch, Barbara (unveröffentlichtes Manuskript): *User expectations towards machine translation: The implications of experience and disconfirmation in neural machine translation*.

Markhardt, Heidemarie (2002): *Das österreichische Deutsch im Rahmen der Europäischen Union das „Protokoll Nr. 10 über die Verwendung österreichischer Ausdrücke der deutschen Sprache“ zum österreichischen EU-Beitrittsvertrag und die Folgen: eine empirische Studie zum österreichischen Deutsch in der EU*. Dissertation, Universität Wien.

Michel, Paul / Neubig, Graham (2018): „Extreme Adaptation for Personalized Neural Machine Translation“, in: *Proceedings of the 56 th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)* 312–318.

Neubarth, Friedrich / Trost, Harald (2017): „Statistische maschinelle Übersetzung vom Standarddeutschen in den Wiener Dialekt“, in: Resch, Claudia / Dressler, Wolfgang U. (Hrsg.): *Digitale Methoden der Korpusforschung*. Wien: Verlag der österreichischen Akademie der Wissenschaften 179–203.

Reynolds, Peter (2015): „Machine translation, translation memory and terminology management“, in: Kockaert, Hendrik J. / Steurs, Frieda (Hrsg.): *Handbook of Terminology*. Amsterdam: John Benjamins Publishing Company 276–287.

Scansani, Randy / Bernardini, Silvia / Ferraresi, Adriano / Gaspari, Federico / Soffritti, Marcello (2017): „Enhancing Machine Translation of Academic Course Catalogues with Terminological Resources“, in: *Proceedings of the Workshop on Human-Informed Translation and Interpreting Technology: Incoma Ltd. Shoumen, Bulgaria* 1–10.

Schmidlin, Regula (2011): *Die Vielfalt des Deutschen: Standard und Variation Gebrauch, Einschätzung und*

Kodifizierung einer plurizentrischen Sprache. Berlin: De Gruyter.

Servan, Christophe / Crego, Josep / Senellart, Jean (2016): *Domain specialization: a post-training domain adaptation for Neural Machine Translation* <https://arxiv.org/pdf/1612.06141>.

Simard, Michel / Ueffing, Nicola / Isabelle, Pierre / Kuhn, Roland (2007): „Rule-based translation with statistical phrase-based post-editing“, in: *Proceedings of the 45 th Annual Meeting of the Association of Computational Linguistics. Prague, Czech Republic — June 23 - 23, 2007* 203–206.

Stolze, Radegundis (2009): *Fachübersetzen - ein Lehrbuch für Theorie und Praxis*. Berlin: Frank & Timme.

Wissik, Tanja (2013): *Terminologische Variation in der Rechts- und Verwaltungssprache eine korpusbasierte Analyse der Hochschulterminologie in den Standardvarietäten des Deutschen in Deutschland, Österreich und der Schweiz*. Dissertation, Universität Wien.

Wloka, Bartholomäus / Budin, Gerhard / Winiwarter, Werner (2013): „Machine Translation, Language Analysis, and Mobile Applications in the Terminology Domain“, in: *Jusletter IT*.