

»textklang« – Ein Mixed-Methods-Workshop zu Lyrik in Text und Ton

Ketschik, Nora

nora.ketschik@ims.uni-stuttgart.de
Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart, Deutschland

Bernhart, Toni

toni.bernhart@ilw.uni-stuttgart.de
Institut für Literaturwissenschaft, Universität Stuttgart, Deutschland

Gärtner, Markus

markus.gaertner@ims.uni-stuttgart.de
Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart, Deutschland

Koch, Julia

julia.koch@ims.uni-stuttgart.de
Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart, Deutschland

Schauffler, Nadja

nadja.schauffler@ims.uni-stuttgart.de
Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart, Deutschland

Kuhn, Jonas

jonas.kuhn@ims.uni-stuttgart.de
Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart, Deutschland

Einleitung

Die Überlieferung von Texten ist vorwiegend an die schriftliche Form gebunden, die bis zur Erfindung von Tonaufnahmetechniken in der zweiten Hälfte des 19. Jahrhunderts die einzige Möglichkeit war, von Sprachbeiträgen nicht nur den Inhalt, sondern weitgehend auch die Darbietung festzuhalten. So haben sich literarische Traditionen und die wissenschaftliche Auseinandersetzung mit Literatur überwiegend entlang der schriftlichen Überlieferung entwickelt. Selbst bei Gattungen wie der Lyrik, in der Klang eine wichtige inhaltliche und ästhetische Rolle spielt (vgl. Richter et al. 2022), steht die Textform im Zentrum der kanonischen Überlieferung. Erst am Ende des 20. Jahrhunderts (u.a. angeregt durch die Sound Studies) haben sich in der Literaturwissen-

schaft Forschungsfelder zu Stimme, Klang, Akustik, Auditivität und Audioliteralität etabliert (Göttert 1998, Meyer-Kalkus 2001, Schulz 2018, Meyer-Kalkus 2020, Meyer-Sieckendiek 2020). Bis auf wenige Ausnahmen (z.B. Rhythmicalizer, vgl. Meyer-Sieckendiek et al. 2017) folgen die Digital Humanities bislang recht stark dieser eingespielten Zugangsweise – obgleich seit etwa 1900 unzählige Tonaufnahmen von Rezitationen vorliegen. Auch in der linguistischen Prosodieforschung und in der Sprachtechnologie wurde über die letzten Jahrzehnte ein Methodeninventar entwickelt, das eine sehr differenzierte Formulierung von Hypothesen zur Beziehung zwischen Text und lautlicher Realisierung erlaubt. Unser Workshop führt empirische Methoden aus der Phonetik mit aktuellen Technologien der Sprachsynthese und literaturwissenschaftlicher Forschung zur Lyrik der Romantik in einem Mixed-Methods-Workflow zusammen und bietet den Teilnehmenden auf diese Weise die Möglichkeit, das Wechselspiel von Textlichkeit und lautlicher Realisierung im Gedichtekorpus explorativ zu erkunden.¹

Der Workshop knüpft an Arbeiten aus dem BMBF-geförderten Projekt »textklang«² an. In »textklang« kooperieren das Deutsche Literaturarchiv (DLA) Marbach sowie das Institut für Maschinelle Sprachverarbeitung und das Institut für Literaturwissenschaft der Universität Stuttgart, die Expertise in unterschiedlichen relevanten Fachgebieten vereinen. Der Fokus des Projekts liegt auf der Erschließung und Analyse lyrischer Texte der Romantik, wobei der Zusammenhang zwischen dem geschriebenen Text und seiner lautlichen Realisierung in Rezitationen und Vertonungen in den Blick genommen wird.

Das beim Workshop verwendete Forschungskorpus zur Lyrik der Romantik speist sich aus der Mediendokumentation des DLA Marbach, die etwa 2700 Audioaufzeichnungen verschiedener Sprecher*innen seit den 1920ern beherbergt. Diese werden im Zuge des Projekts digitalisiert und um die dazugehörigen Metadaten und Transkripte ergänzt; darüber hinaus werden Texte und Rezitationen mit automatisch erzeugten Annotationen angereichert (siehe Schauffler et al. 2022b für eine Übersicht). Aktuell umfasst das »textklang«-Korpus 1261 Audioaufnahmen zu 786 Gedichten. Metadaten, Textdateien und lizenzfreie Audiodaten werden kontinuierlich über eine interaktive Webseite veröffentlicht.³

In unserem Workshop kommen alle Bereiche des Mixed-Methods-Workflows zum Einsatz, indem Ansätze aus traditionell sehr unterschiedlich arbeitenden Disziplinen zusammengeführt werden. Das Analysetool ICARUS (Gärtner et al. 2015) unterstützt den korpus- und textorientierten Zugang, bildet dabei aber neben morphosyntaktischen Annotationen der Texte auch die phonetischen Annotationen der Rezitationen ab. Hierfür kommen Verfahren aus der Phonetik zum Einsatz, die die Eigenschaften des Sprachsignals systematisch erfassen. Sprachtechnologische Verfahren der Signalanalyse und -manipulation ermöglichen es so dann, bestimmte Annahmen über ein Re-Synthese-Tool kontrolliert zu testen. Der Bedarf für ein so weit gefasstes Methodenspektrum folgt aus den Grundeigenschaften des Untersuchungsgegenstands selbst. Der Workshop leistet einen Beitrag, die fachspezifischen Ansätze methodologisch zusammenzuführen und auf diese Weise

den insbesondere für Lyrik zentralen Zusammenhang von Text und Klang in den Blick zu rücken.

Use-Cases

Idee des Workshops ist, dass die Teilnehmenden ihre eigenen Fragestellungen an Rezitationen von Lyrik der Romantik mitbringen können und darauf aufbauend während der Datenexploration Hypothesen entwickeln. Alternativ können die von uns vorgeschlagenen Fragestellungen aufgegriffen werden. Im Workshop thematisieren wir mehrere Use-Cases aus dem Projektkontext, darunter die Realisierung paralleler Strukturen (z.B. Reim, Satzbau), die unter strukturellen, semantischen und melodischen Aspekten von Interesse sind. Eine andere Fallstudie untersucht unterschiedliche Realisierungen von Enjambements (Schauffler et al. 2022a), die im Spannungsfeld von Vers- und Satzstruktur stehen. In Rezitationen können Sprecher*innen die syntaktische Einheit betonen, die Versgrenze markieren oder einen Mittelweg wählen (vgl. Tsur und Gafni 2019).

Ein weiterer Anwendungsfall, der exemplarisch etwas näher erläutert werden soll, beschäftigt sich mit Interjektionen. Interjektionen bezeichnen Ausrufe- oder Empfindungsworte (z. B. „ach“, „oh“, „juchhe“) und stehen im Grenzbereich von Schriftlichkeit und Mündlichkeit (Wharton 2003, Liedtke 2019). Sie nehmen eine syntaktische Sonderrolle ein und werden in der Linguistik als eigenständige Klasse behandelt, den Partikeln zugeordnet oder als Satzäquivalente angesehen (Liedtke 2019). Sie tragen einerseits denotativ keine Bedeutung, bringen andererseits Emotionen verschiedenster Art und in unterschiedlichen Intensitätsgraden zum Ausdruck (Schwarz-Friesel 2013, 155-157). Mit dem hier vorgestellten Mixed-Methods-Ansatz soll der Spielraum und der besondere textlich-klangliche (Zwischen-)Status von Interjektionen untersucht werden. Dabei interessiert zum einen die syntaktische Stellung von Interjektionen, zum anderen ihr Bedeutungsspektrum sowie, als dritter Aspekt, ihre lautliche Ausprägung. Die „Offenheit“ dieser Wortart legt die Hypothese nahe, dass die verschiedenen Ebenen sich gegenseitig beeinflussen können, beispielsweise das syntaktische Umfeld die lautlichen Realisierungen in der Rezitation prägt oder bestimmte klangliche Merkmale die Bedeutung von Interjektionen ausmachen.

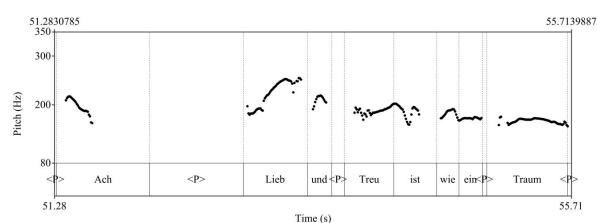


Abb.1

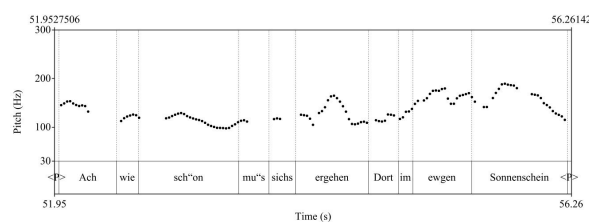


Abb.2

Die abgedruckten Beispiele deuten die syntaktisch-lautlichen Spielräume der Interjektion „Ach“ im Gedichtekorpus an: Während sie im ersten Beispiel syntaktisch isoliert steht (markiert durch den Tonhöhenverlauf und die Sprechpause), wird sie im zweiten Beispiel syntaktisch und lautlich in den Satz integriert. Auch die mit dem „Ach“ ausgedrückten Emotionen (im ersten Beispiel Schmerz, im zweiten Freude) changieren und werden – neben dem semantischen Kontext des Wortes – von der jeweiligen sprachlichen Realisierung beeinflusst. Mögliche Leitfragen für weitere Untersuchungen könnten sein: Welche syntaktischen Merkmale von Interjektionen gehen mit welchen lautlichen Merkmalen einher? Werden Interjektionen in gleicher (syntaktischer) Position lautlich parallel realisiert? Welche Varianz ist zwischen unterschiedlichen Sprecher*innen zu beobachten? Inwiefern beeinflusst die lautliche Realisierung die Bedeutung oder Wahrnehmung von Interjektionen?

Tools

Icarus

Für die Exploration und Visualisierung des Korpus mit allen Annotationsebenen verwenden wir ICARUS (Gärtner 2015) als Anfrageschnittstelle. ICARUS erlaubt eine gemeinsame Visualisierung von prosodischen Informationen und klassischen morphosyntaktischen Annotationen. Darüber hinaus können gezielt Anfragen unter Einbeziehung aller im Korpus verfügbaren Annotationsebenen gestellt werden, um Instanzen bestimmter Phänomene zu finden. An Annotationen stehen sämtliche für das GRAIN Korpus (Schweitzer et al. 2018) beschriebenen morphosyntaktischen und prosodischen Ebenen zur Verfügung. Darüber hinaus sind die Gedichte auch mit Markierungen zu Vers- und Strophenenden versehen, welche ebenfalls in Abfragen benutzt werden können. Je nach Entwicklungsfortschritt wird ICARUS als Desktop-Applikation⁴ eingesetzt oder in der Variante einer auf das »textklang«-Korpus zugeschnittenen Web-Oberfläche bereitgestellt.

IMS Speech Synthesis Toolkit Toucan

Die durch die Datenexploration entwickelten Hypothesen über Zusammenhänge zwischen Text und lautsprachlicher Realisierung sollen in Perzeptionsexperimenten untersucht werden. Mittels Sprachsynthese erstellen wir zu diesem Zweck eine prosodische Repli-

kation der Originalaufnahmen, wobei phonetische Details (z.B. Lautdauer, Tonhöhe) gezielt manipuliert werden können (Koch et al. 2022). Unser Synthesemodell basiert auf der Modellarchitektur von FastSpeech 2 (vgl. Ren 2021), für die Implementierung nutzen wir das open-source Toolkit IMS Toucan⁵ (Lux et al. 2021, Lux und Vu 2022). Die Workshopteilnehmer*innen können über eine Bedienoberfläche mit dem Modell interagieren, indem sie spezifische, mit einem Phänomen verbundene Merkmale verändern und anschließend die Effekte der veränderten Parameter in der Perzeption testen. Beispielsweise kann die Länge, mit der ein Sprecher etwa das Versende markiert, verkürzt werden, die Tonhöhe an einer bestimmten Stelle angepasst oder die Dauer von Pausen verändert werden.

Ablauf und Ziele

Wir beginnen den Workshop mit einer Einführung in den multimodalen Ansatz und adressieren die methodologisch wie wissenschaftstheoretisch relevante Frage, wie die Spezialisierungen der Fachgebiete innerhalb der DH sinnvoll zusammengeführt werden können. Anschließend präsentieren wir mögliche Forschungsbeispiele und führen in die verwendeten Tools ein.

In zwei Praxisrunden haben die Teilnehmenden die Möglichkeit, das Lyrikkorpus zu erforschen, eigene Forschungsfragen zu entwickeln sowie diese exemplarisch zu untersuchen. Dies kann individuell oder in Kleingruppen geschehen. Die erste Praxisrunde dient der Exploration des Korpus und der Entwicklung möglicher Hypothesen. Hierfür kommt das Tool ICARUS zum Einsatz, über das die Teilnehmer*innen die verschiedenen Annotatensebenen (u.a. morphosyntaktisch, phonetisch) sichten und komplexe Suchanfragen an die Texte modellieren können. Auf Grundlage der Annotationen zur Text- und Lautgestalt können Forschungsfragen entwickelt oder eine der vorgestellten Fragestellungen aus der theoretischen Einführung exploriert werden. Nach einer Zusammenschau der Hypothesen dient die zweite Praxisrunde dazu, ausgewählte Fragestellungen probeweise zu validieren, indem die Annahmen in das Sprachsynthesemodell überführt werden. Wenn beispielsweise die Annahme besteht, dass die Länge und die Tonhöhe einen Einfluss darauf haben, ob die "bedeutungsfreie" Interjektion "Ach" negativ oder positiv konnotiert ist, können eben diese Merkmale in der Sprachsynthese gezielt modifiziert und die Effekte dieser Veränderungen getestet werden.

Die Ziele des Workshops bestehen folglich darin, die Möglichkeiten des Mixed-Methods-Ansatzes auszuschöpfen und Lyrik in ihrer Multimodalität erforschbar zu machen. Dabei liegt ein besonderer Schwerpunkt darauf, zu zeigen, wie fruchtbar das Zusammenspiel von textlicher und klanglicher Ebene sein kann. Zwar können die zu behandelnden Fragestellungen im Rahmen des Workshops nur ansatzweise durchgespielt werden, sie können dabei aber die Potenziale des interdisziplinären Ansatzes offenlegen.

Anhang

Zeitplan

1. Einführung und Ablauf (15 Min)
2. Theoretischer Teil (30 Min)
 1. Vorstellung der Projektidee
 2. Einführung in die Use-Cases
 3. Einführung in die verwendeten Tools (anschließende Pause, 15 Min)
3. Praktischer Teil
 1. Erste Praxisrunde: Exploration der Daten, Entwicklung von Hypothesen (45 Min)
 2. Sammeln der Ergebnisse, Vorstellung möglicher Fragestellungen (15 Min) (anschließende Pause, 30 Min)
 3. Zweite Praxisrunde: Bearbeitung der Fragestellungen, Syntheseexperimente (45 Min)
 4. Sammeln der Ergebnisse (15 Min)
4. Abschlussdiskussion (30 Min)

Teilnehmer*innen

Unser Workshop ist für ca. 20 Teilnehmer*innen geeignet und richtet sich an Interessierte aus den digitalen Geisteswissenschaften. Bestimmte technische Vorkenntnisse sind nicht erforderlich.

Technische Ausstattung

Die Teilnehmenden arbeiten an ihren eigenen Laptops. Ausreichend Steckdosen, stabiles Wifi und ein Beamer sollten vorhanden sein. Installationshinweise werden im Vorfeld an die Teilnehmer*innen verschickt.

Beitragende

Nora Ketschik (Institut für Maschinelle Sprachverarbeitung (IMS), Universität Stuttgart, nora.ketschik@ims.uni-stuttgart.de) ist wissenschaftliche Mitarbeiterin an der Universität Stuttgart. Sie promoviert zu Netzwerkanalysen von mittelhochdeutschen Romanen und setzt sich kritisch mit der Verwendung computergestützter Methoden für literaturwissenschaftliche Analysezwecke auseinander.

Toni Bernhart (Institut für Literaturwissenschaft, Universität Stuttgart, toni.bernhart@ilw.uni-stuttgart.de) ist Privatdozent für Neuere deutsche Literatur und wissenschaftlicher Mitarbeiter der Abteilung Digital Humanities an der Universität Stuttgart. Seine Forschungsschwerpunkte sind die Imaginationsgeschichte von 'Volks poesie', Auditivität und Literatur, Quantitative Literaturwissenschaft und Wissenschaftsgeschichte der Digital Humanities.

Markus Gärtner (IMS, Universität Stuttgart, markus.gaertner@ims.uni-stuttgart.de) ist wissenschaftlicher Mitarbeiter und Doktorand an der Universität Stuttgart und regelmäßig in der technischen Konzeption und Umsetzung von infrastrukturell fokussierten Projekten tätig.

Julia Koch (IMS, Universität Stuttgart, julia.koch@ims.uni-stuttgart.de) ist wissenschaftliche Mitarbeiterin und Doktorandin an der Universität Stuttgart. In ihrer Promotion arbeitet sie an Deep Learning Modellen für Sprachsynthese mit besonderem Fokus auf Kontrollierbarkeit.

Nadja Schauffler (IMS, Universität Stuttgart, nadja.schauffler@ims.uni-stuttgart.de) ist wissenschaftliche Mitarbeiterin an den Instituten für Maschinelle Sprachverarbeitung und Linguistik an der Universität Stuttgart und Postdoc im Projekt »textklang«, wo sie sich vor allem mit prosodischer Varianz beschäftigt.

Jonas Kuhn (IMS, Universität Stuttgart, jonas.kuhn@ims.uni-stuttgart.de) ist Professor für Computerlinguistik am Institut für Maschinelle Sprachverarbeitung und seit vielen Jahren an interdisziplinären Projekten zur Methodenentwicklung für die Digital Humanities beteiligt. Er ist federführender Projektleiter des BMBF-Projekts »textklang«.

Fußnoten

1. Rollen der Beitragenden: Nora Ketschik (Writing - original draft, Investigation, Methodology), Toni Bernhart (Writing - review and editing), Markus Gärtner (Software), Julia Koch (Software), Nadja Schauffler (Writing - original draft, Investigation, Methodology), Jonas Kuhn (Conceptualization, Methodology, Supervision).
2. <https://textklang.org/> (19.07.2022).
3. Interaktive Übersicht des »textklang«-Korpus: <https://clarin03.ims.uni-stuttgart.de/keshif/demo/textklang.html> (19.07.2022). Die Übersicht ist auch unter "Data" auf der Projektseite (<https://textklang.org/>) abrufbar.
4. ICARUS ist unter <https://github.com/ICARUS-tooling/icarus1-platform> (19.07.2022) bereits open source verfügbar und kann im Voraus von Teilnehmer*innen heruntergeladen werden.
5. <https://github.com/DigitalPhonetics/IMS-Toucan> (19.07.2022).

Bibliographie

Gärtner, Markus, Katrin Schweitzer, Kerstin Eckart und Jonas Kuhn. 2015. "Multi-modal Visualization and Search for Text and Prosody Annotations." In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing: System Demonstrations*.

Göttert, Karl-Heinz. 1998. *Geschichte der Stimme*. München: Fink.

Koch, Julia, Florian Lux, Nadja Schauffler, Toni Bernhart, Felix Dieterle, Jonas Kuhn, Sandra Richter, Gabriel Viehhauser und Ngoc Thang Vu. 2022. "PoeticTTS - Controllable Poetry Reading for Literary Studies." In *Proceedings of Interspeech 2022*.

Liedtke, Frank und Lena Rosenbaum. 2019. "Interjektionen und Kontextbezug. Pragmatische Templates als Analysemodell." In *Expressivität im Deutschen*, hg. von

Franz d'Avis und Rita Finkbeiner, 129-148. Berlin/Boston: De Gruyter. 10.1515/9783110630190.

Lux, Florian, Julia Koch, Antje Schweitzer und Ngoc Thang Vu. 2021. "The IMS Toucan system for the Blizzard Challenge 2021." In *Proceedings of the Blizzard Challenge Workshop*.

Lux, Florian und Thang Vu. 2022. "Language-Agnostic Meta-Learning for Low-Resource Text-to-Speech with Articulatory Features." In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics*.

Meyer-Kalkus, Reinhart. 2001. *Stimme und Sprechkünste im 20. Jahrhundert*. Berlin: Akademie Verlag.

Meyer-Kalkus, Reinhart. 2020. *Geschichte der literarischen Vortragskunst*. Berlin: Metzler. <https://doi.org/10.1007/978-3-476-04802-8>.

Meyer-Sickendiek, Burkhard. 2020. *Hörlyrik. Eine interaktive Gattungstheorie*. Paderborn: Fink.

Meyer-Sickendiek, Burkhard, Hussein Hussein und Timo Baumann. 2017. „Rhythmicalizer. Data Analysis for the Identification of Rhythmic Patterns in Readout Poetry." In *INFORMATIK 2017. Lecture Notes in Informatics (LNI) - Proceedings*, hg. von Maximilian Eibl und Martin Gaedke, 2189-2200. Bonn: Köllen Druck + Verlag GmbH (Series of the Gesellschaft für Informatik 275).

Ren, Yi, Chenxu Hu, Xu Tan, Tao Qin, Sheng Zhao, Zhou Zhao und Tie-Yan Liu. 2021. "FastSpeech 2: Fast and High-Quality End-to-End Text to Speech." In *International Conference on Learning Representations*.

Richter, Sandra, Toni Bernhart, Felix Dieterle, Gabriel Viehhauser, Gunilla Eschenbach, Jonas Kuhn, Nadja Schauffler, André Blessing, Markus Gärtner, Kerstin Jung, Nora Ketschik, Anna Kinder, Julia Koch, Thang Vu und Andreas Kozlik. 2022. "Der Klang der Lyrik. Zur Konzeptualisierung von Sprecher und Stimme, auch für die computationelle Analyse." *Poema. Jahrbuch für Lyrikforschung / Annual for the Study of Lyrical Poetry / La recherche annuelle en poésie lyrique* 1 (im Erscheinen).

Schauffler, Nadja, Fabian Schubö, Toni Bernhart, Gunilla Eschenbach, Julia Koch, Sandra Richter, Gabriel Viehhauser, Thang Vu, Lorenz Wesemann und Jonas Kuhn. 2022a. "Prosodic realisation of enjambment in recitations of German poetry." In *Proceedings of the 11th international Conference on Speech Prosody*, 530-534. 10.21437/SpeechProsody.2022-108

Schauffler, Nadja, Toni Bernhart, André Blessing, Gunilla Eschenbach, Markus Gärtner, Kerstin Jung, Anna Kinder, Julia Koch, Sandra Richter, Gabriel Viehhauser, Thang Vu, Lorenz Wesemann und Jonas Kuhn. 2022b. "»textklang« - Towards a Multi-Modal Exploration Platform for German Poetry." In *Proceedings of the 13th edition of the Language Resources and Evaluation Conference (LREC)*, 5345-5355.

Schulz, Miklas. 2018. *Hören als Praxis. Sinnliche Wahrnehmungsweisen technisch (re-)produzierter Sprache*. Wiesbaden: Springer (Auditive Vergesellschaftungen Hörsinn - Audiotechnik - Musikerleben). <https://doi.org/10.1007/978-3-658-19654-7>.

Schwarz-Friesel, Monika. 2013. *Sprache und Emotion*. 2. Aufl. Tübingen: Narr Francke Attempto Verlag.

Schweitzer, Katrin, Kerstin Eckart, Markus Gärtner, Agnieszka Falenska, Arndt Riester, Ina Rösiger, Antje Schweitzer, Sabrina Stehwein und Jonas Kuhn. 2018. "German Radio Interviews: The GRAIN Release of the SFB732 Silver Standard Collection." In *Proceedings of*

the 11th edition of the Language Resources and Evaluation Conference (LREC).

Tsur, Reuven und Chen Gafni. 2019. "Enjambment - irony, wit, emotion. A case study suggesting wider principles." *Studia Metrica et Poetica* (5): 7-28.

Wharton, Tim. 2003. "Interjections, Language, and the 'Showing/Saying' Continuum." *Pragmatics and Cognition* 11(1): 39-91. [10.1075/pc.11.1.04wha](https://doi.org/10.1075/pc.11.1.04wha) .