

# A Taxonomy for Deep Learning in Natural Language Processing

Severin Landolt  
University of St.Gallen  
[sev.landolt@gmail.com](mailto:sev.landolt@gmail.com)

Thiemo Wambsganß  
University of St.Gallen  
[thiemo.wambsganss@unisg.ch](mailto:thiemo.wambsganss@unisg.ch)

Matthias Söllner  
University of Kassel  
[soellner@uni-kassel.de](mailto:soellner@uni-kassel.de)

## Abstract

*Despite a large number of available techniques around Deep Learning in Natural Language Processing (NLP), no holistic framework exists which supports researchers and practitioners to organise knowledge when designing, comparing and evaluating NLP applications. This paper addresses this lack of a holistic framework by developing a taxonomy for Deep Learning in Natural Language Processing. Based on a systematic literature review as proposed by Webster and Watson [1] and vom Brocke et al. [2] and the iterative taxonomy development process of Nickerson et al. [3] we derived five novel dimensions and 38 characteristics based on a sample of 205 papers. Our research suggests, that a Deep Learning NLP approach can be distinguished by five dimensions which were partly derived from the CRISP-DM methodology: application understanding, data preparation, modeling, learning technique and evaluation. We, therefore, hope to provide guidance and support for researchers and practitioners when using Deep Learning for NLP to design, compare and evaluate NLP applications.*

## 1. Introduction

Back in 2017, Gereon Frahling, founder of DeepL said in an interview: “Five years ago I would certainly have said that we don’t do full-text machine translation, it’s too complicated for us”, recalls Frahling. “Google had developed its old system over ten years with a lot of people and messing with it just didn’t make sense” [4]. This quote has been translated from German into English by solely using the DeepL Translator. It reflects the popularity and extensive research in the field of *Artificial Intelligence* (AI). One dynamic field of AI is *Deep Learning* (DL). DL emerged as especially promising for solving complex problems (e.g. machine translation or computer vision) and was primarily developed over the last ten years [5, 6, 7].

This development has led to a significant increase

in novel approaches for DL in NLP. Moreover, the combination of NLP and DL has been successfully embedded in user-centered artefacts, e.g., to enhance individual learning of students with adaptive writing support systems [8] or adaptive tutoring through a smart personal assistant [9]. A brief search reveals that there are over 40 different applications of NLP using a neural architecture [10]. However, due to rapid growth of approaches and techniques of DL in NLP, current research lacks a holistic view on different concepts of applying the new technology for researchers to design, evaluate and compare NLP studies.

Taxonomies are a usual approach in Information Systems research to summarise the current state of knowledge of a novel technology, e.g., for the design of conversational agents [11, 12] or text mining features [13]. Hence, we aim to develop a taxonomy through a conceptual and empirical investigation of NLP literature and research. Our paper contributes to the literature on text mining and DP for NLP by answering the following research question:

**RQ:** *What are the theoretically grounded and empirically validated key components of Deep Neural Networks used in Natural Language Processing?*

In order to contribute to our research question, we: 1) conducted a systematic literature review as proposed by Webster and Watson [1] and vom Brocke et al. [2] to investigate the manifold techniques of DL in NLP based on a sample of 205 papers; 2) analysed the different concepts and models and derived a novel taxonomy with five distinct dimensions and 38 characteristics following the taxonomy development process of Nickerson et al. [3]; and 3) provide an example of how the taxonomy can be applied by fellow researchers and practitioners in order to offer a set of techniques for them to design, compare and evaluate existing NLP challenges and applications.

With this paper, we want to encourage researchers to contribute to the existing taxonomy and add missing dimensions or characteristics, so that the resulting taxonomy supports DL and NLP researchers and

practitioners in A) identifying and organising the components which may be combined in the creation of NLP models and B) in systematically describing how NLP models are configured to perform different tasks.

To support researchers, we have developed a holistic framework (taxonomy) which is empirically grounded. Furthermore, the taxonomy is complemented by drawing knowledge from the established Cross Industry Standard Process for Data mining (CRISP-DM) methodology. This taxonomy should assist as a tool for the design, comparison and evaluation by classifying different approaches of NLP using DL into their integral parts.

The remainder of this paper is structured as follows: first, we introduce the reader to the conceptual background of Deep Learning and Natural Language Processing. Next, we briefly outline the related work on holistic DL NLP framework and thereby, illustrated the research gap we aim to address. Then, we sketch our followed research methodology, the taxonomy development process and conclude by discussing the results.

## 2. Conceptual Background and Related Work

In this section, we outline the conceptual background of DL and NLP. Moreover, we illustrate the related work on holistic DL frameworks for NLP and provide an overview of potential research gaps.

### 2.1. Neural Networks and Deep Learning

Artificial Neural Networks (ANN), as the name already reveals, are computational networks that are able to solve complex, nonlinear mathematical problems [6]. The field of ANN has been inspired by the ambition to model biological neural systems. NN are modelled as collections of layers of neurons that are connected in an acyclic graph [14]. The output of such ANN could be a predicted numerical value, but in many cases, it is usually taken to represent the class scores (e.g. in text classification) [15]. In most cases of NLP, we are interested in multinomial classification, such as part-of-speech tagging. Hence, the output layer yields a probability distribution across the output nodes. Goldberg describes DL as: "... a branch of machine learning. It is a re-branded name for neural networks—a family of learning techniques that were historically inspired by the way computation works in the brain" [16, p. 7]. DNNs stack up several hidden layers, with each layer acting as the input to the next layer. DL allows a computer to build complex concepts out of simpler concepts [17], [18]. Another perspective on deep

learning is that the depth allows the computer to learn a multi-step program [18]. Each layer can be interpreted as the state of the computer's memory after executing a set of instructions. We provide a non-exhaustive high-level overview of different DL techniques [19].

- *Multilayer Perceptron (MLP)* is a feed-forward neural network with multiple (one or more) hidden layers between the input layer and output layer.
- *Autoencoder (AE)* is an unsupervised model attempting to reconstruct its input data in the output layer.
- *Convolutional Neural Network (CNN)* is a special kind of feed-forward neural network with convolution layers and pooling operations.
- *Recurrent Neural Network (RNN)* use loops and memories in to remember former computations.
- *Deep Reinforcement Learning (DRL)*. Reinforcement learning operates on a trial-and-error paradigm. The whole framework mainly consists of the following components: agents, environments, states, actions, and rewards.

### 2.2. Natural Language Processing

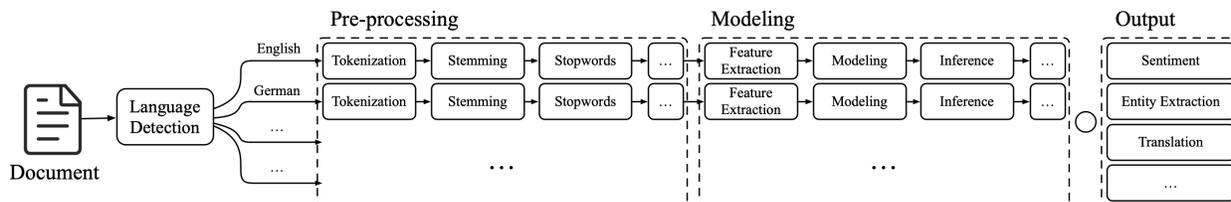
Over the last decade, the increase in available textual data and use of DNNs has propelled forward the field of NLP. What once used to be just a bag-of-words approach has become a major discipline in several academic fields such as computer science, psychology or linguistics. Otter et al. broadly define NLP as: "... the engineering of computational models and processes to solve practical problems in understanding human languages" [20, p. 2].

Figure 1 depicts a generalised comparison between a classical and deep learning-based NLP workflow. Using a classical approach, the data is first pre-processed which usually includes tokenization, pruning, stopword removal, filtering or stemming [22, 23]. Then, features are generated and used as an input to train a model. Depending on the application, the output can be the sentiments, translations, information retrieval, etc.).

In a DL workflow, the document is first pre-processed. The resulting artefact is then processed by a word embedding algorithm (e.g. *Word2Vec*) and passed over to one or more NN(s).

A core element of NLP in DL are *word embeddings*. Goldberg describes word embeddings as follows: "A major component in neural networks for language is

## Classical NLP



## Deep Learning-based NLP

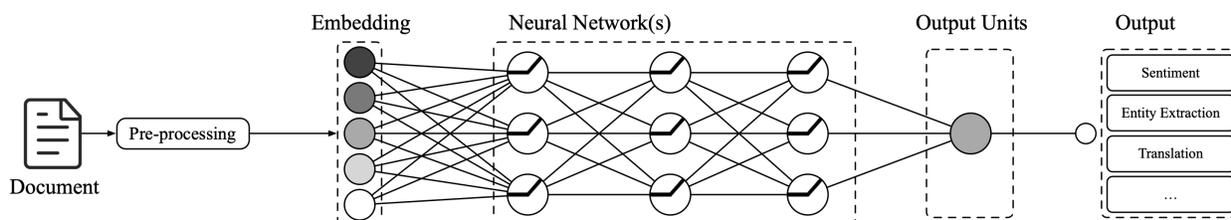


Figure 1. A comparison of classical NLP versus DL oriented NLP process. [21]

the use of an embedding layer, a mapping of discrete symbols to continuous vectors in a relatively low dimensional space” [16, p. 3]. This way of representing context as embeddings, allows language models to generalise unseen data very well [15]. When embedding words, they are transformed from isolated, distinct symbols into mathematical objects that can be operated on [16]. Word embeddings are generated by using already pre-trained neural nets/algorithms, where its input is a text corpus. An early and popular embedding algorithm is known as *Word2Vec* and was developed by Tomas Mikolov from Google [24, 16]. Today, there exist many variants of the *Word2Vec* algorithm such as *fastText* (<https://fasttext.cc>) or *GloVe* [25].

Otter et al. [20] underline the importance of gained understanding of real-world problems, and that a pure engineering perspective is meaningless in and of itself. Given the research interest in NLP, there are several different NLP applications. For instance, Ruder compiled a non-exhaustive list of over 40 NLP applications [10]. From this list as well as the deep learning NLP workflow, it can be derived, that there is a need for a holistic framework in order to categorise those applications and how they fit together with the configuration and workflow of DNNs.

### 2.3. Holistic Frameworks for DL in NLP

We explain the scientific relevance of this paper in three steps. First, there is an exceptionally active research community and industry interest. Recent breakthroughs in the field of DL and Natural Language Processing NLP show significant interest. The interest

of the research community is reflected in the number of academic papers published. Since 2010, the number of submitted papers in computer science has risen six-fold to 34'000 papers per year [26]. Given the rapid and successful development and emergence of new DL models and NLP applications (e.g., for argumentation mining [27]), it is vital to understand the similarities and differences among different models. As mentioned in the introduction, taxonomies provide a common way in Information Systems research to summarise the current state of a technology [13, 11, 28, 29, 30, 31].

Secondly, despite the significant rise of interest and extensive research in the field of AI in the past decade, there is no comprehensive taxonomy that includes DL and NLP. As of today, there exist only a handful of simple, non-systematic approaches [32, 33, 34]. One type of taxonomic scheme only organises types of Neural Networks (NN) with respect to their architecture. For example, Gardner & Dorling present such an approach where NN are clustered into a hierarchy of two groups [32]. Another approach to a framework are matrices. For instance, Gollapudi structures DL into four areas, using one axis for learning type (supervised/unsupervised) and the other axis to differentiate between shallow and deep learning [33].

A third approach used by Perez’ ”The deep learning AI playbook” is to pick different algorithms for DL and describe them with their corresponding elements [34]. A similar approach was taken by Ruder to describe transfer learning [35]. While these taxonomies and frameworks provide valuable insights for the design and understanding of NN or DL, in this paper two relevant gaps in the current literature were identified.

## 2.4. Research Gaps

Firstly, the approaches from the previous section are not sufficiently comprehensive because they miss a transparent and systematic review process. This makes it harder for researchers and practitioners to derive the comprehensiveness and thus the quality of a framework. Hence, the first gap can be described as a *lack of a systematic review process*. In our taxonomy we try to identify  $j$  characteristics for  $i$  dimensions, using established frameworks for both, literature review [1, 2] and taxonomy development [3].

Secondly, the existing frameworks highlight one specific aspect and often have a stark focus on the technical representation of DL in NLP [36, 37, 38]. Conversely, applications of DL in NLP have many different components (e.g. embedding or performance measures). In order to derive crucial elements to fully describe DL in NLP, a more holistic approach has to be taken. This gap is about the *exhaustiveness of identified characteristics and dimensions*. This claim is supported by several authors who point out that the actual structuring of the existing models has been insufficiently investigated [39, 40]. Ferilli et al. [39] confirm demand for a taxonomy, arguing that the relationships underlying DL call for some kind of taxonomic background.

## 3. Research Methodology

In this paper, the components of NLP models are identified and organised using the method for taxonomy development suggested by Nickerson et al. [3]. The term taxonomy can be seen as a synonym of “typology”, “framework” or “classification”. Taxonomies are useful ways of representing knowledge about objects in a domain [41]. This is done by classifying objects of interest. Hence, taxonomies can help explain similarities and differences among described objects. A taxonomy can be formally described as follows:

$$T = \{D_i, i = 1, \dots, n \mid D_i = \{C_{ij}, j = 1, \dots, k_i, k_i \geq 2\}\}$$

where

$D_i = \text{dimension } i$

$C_{ij} = \text{characteristic } j \text{ for dimension } i$  such that  $C_{ij}$  ( $j = 1, \dots, k_i$ ) are mutually exclusive and collectively exhaustive for *dimension*  $i$ . The mutual exclusive restriction means that no object can have two different characteristics in a dimension and collectively exhaustive is used when each object has at least one characteristic in each dimension [3, 6]. The development method itself is iterative and includes the following steps:

1. Determine a meta-characteristic: Choosing a *meta-characteristic* limits the problem of “naïve empiricism” in which a large number of characteristics are defined, hoping that a useful pattern will emerge. This meta-characteristic is the base of all other characteristics of the final taxonomy, where each characteristic should be a logical consequence of the meta-characteristic [3].
2. Determine ending conditions: It is important to mention that despite the formal representation of taxonomies, it is not the goal to develop “best” or “correct” taxonomies as they cannot be defined [3]. The process attempts to discover “effective–or–useful solutions” [3, 42]. Given this lack of metrics, Nickerson et al. propose five qualitative attributes that a useful taxonomy should comply with [3]. Those attributes also represent the *subjective ending conditions* in the process of taxonomy development.
  - Conciseness
  - Robustness
  - Comprehensiveness
  - Extendibility
  - Explanatory

In addition, there are eight *objective ending conditions* [3]:

- All objects (or a representative sample) were analysed
- No object was merged or split in the last iteration
- At least one object assigned to each characteristic
- No new dimensions or characteristics were added in the last iteration
- No dimensions or characteristics were merged or split in the last iteration
- Every characteristic within the dimension is unique
- Every dimension is unique
- Every combination of characteristics is unique

The final taxonomy should fulfil all ending conditions (subjective and objective) as well as the initial definition of a taxonomy (i.e. mutually exclusive and collectively exhaustive).

3. Select an empirical-to-conceptual or conceptual-to-empirical approach: The iterative process

is at the core of the taxonomy development method [3]. After the meta-characteristic and the ending conditions are determined, the iteration process starts. One can start with either an *empirical-to-conceptual* or *conceptual-to-empirical* approach [3].

In the empirical-to-conceptual approach (see 4e–6e in Figure 2), focus lies on deriving characteristics and dimensions based on the identification of common characteristics from a sample of objects (e.g. literature review).

The second approach is conceptual-to-empirical (see 4c–6c in Figure 2). This deductive process starts by using extant theory to derive characteristics and dimensions.

4. Iteratively follow the approach, until the ending conditions are met, which ends the iteration: After each iteration of the taxonomy, the satisfaction of all ending conditions from are checked. If all the ending conditions are met, the development process is completed. It occurs that some dimensions or characteristics are eliminated or combined so that fewer dimensions and/or characteristics result in the taxonomy [3].

Figure 2 summarises the taxonomy development method suggested by Nickerson et al. [3].

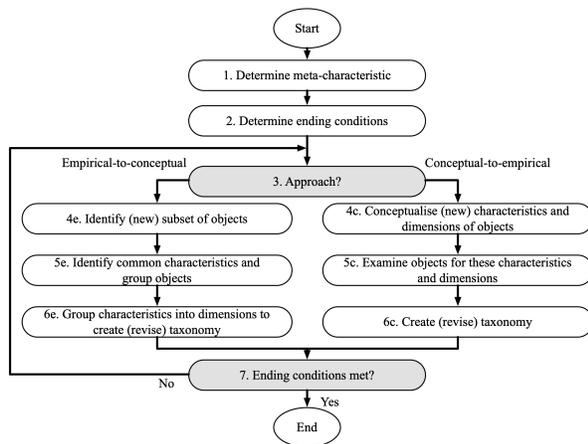


Figure 2. Taxonomy development method. Adapted from [3]

## 4. Taxonomy Development

### 4.1. Systematic Literature Review

To clearly define the scope of a literature review and the resulting taxonomy, this paper draws on an

established framework for literature reviews presented by Cooper [43]. The characteristics of the review are summarised in Table 1. Note that a particular review can employ more than one strategy within a characteristic [43].

Table 1. A taxonomy of literature reviews. Adapted from [43]

Characteristic		Categories			
1	focus	research outcomes	research methods	theories	applications
2	goal	integration		criticism	central issues
3	perspective	neutral representation		espousal of position	
4	coverage	exhaustive	exhaustive & selective	representative	central/ pivotal
5	organisation	historical		conceptual	methodological
6	audience	specialised scholars	general scholars	practitioners	general public

Webster and Watson [1] and vom Brocke et al. [2] make clear, that the process of including and excluding sources has to be as transparent as possible to ensure the resulting artefact is credible. Table 2 outlines the accessed databases, review type, the exact search string(s) used, covered time period, the resulting number of hits and relevant papers for each database and cross-references. A paper was marked as “relevant” if it used and described applied NLP (cf. Table 1: (1) focus: applications). The search strings were constructed around the key terms *deep learning* [18, 6] and *natural language processing* [17, 16, 44, 20]. The reason for the four different search strings lies in the different keyword mechanisms of the databases. In order to obtain sufficient and relevant academic literature, we adjusted the search strings accordingly to the database.

In addition, we obtained a second sample by looking at cross-references. Webster and Watson [1] call this procedure *Go Backward*. It includes reviewing the citations for the articles already identified. The aim is to find previous articles that should be considered. A total of 205 papers (two duplicates have been removed from initially 207 papers) ranging from 1997–2019 have been reviewed. We used these for the empirical-to-conceptual iterations.

### 4.2. Iterative Taxonomy Development

The meta-characteristic for this paper is defined to be the *key components of Deep Neural Networks used in Natural Language Processing*. The second step is the determination of the ending conditions. We use the following the ending conditions to determine when to terminate the iterative process: Definitions of a

**Table 2. Panel A: Considered sources. Panel B: Used search strings. Data retrieved in May 2019 \*IEEE’s default search including abstract, title text, and indexing terms.**

Panel A						
Name	Source Type	Review Type	Search String	Coverage	Hits	Relevant
Cross-references	Mixed	Mixed	–	1997–2018	71	71
arXiv.org	Preprint Repository	Moderation	See Panel B: 1	1997–2019	939	63
IEEE	Database	Peer-Review	See Panel B: 2	1997–2019	671	38
Science Direct	Database	Peer-Review	See Panel B: 3	1997–2019	1501	20
AIS eLibrary	Database	Peer-Review	See Panel B: 4	1997–2019	446	15
Duplicates Removed	–	–	–	–	–	2
<b>Total</b>	–	–	–	–	3628	205

Panel B		
String No.	Search String	Search Field
1	("deep learning" OR "neural networks") AND (nlp OR "natural language processing" OR nlu OR "natural language understanding" OR "text mining" OR "human language" OR "information retrieval" OR "ontology" OR "opinion mining")	"all fields"
2	("deep learning") AND ("nlp" OR "nlu" OR "natural language processing" OR "natural language understanding")	"metadata only"*
3	("deep learning") AND (nlp OR "natural language processing" OR nlu OR "natural language understanding" OR "text mining" OR "text analysis")	"all fields"
4	("deep learning" OR "neural networks") AND (nlp OR "natural language processing" OR nlu OR "natural language understanding" OR "text mining" OR "human language" OR "information retrieval" OR "ontology" OR "opinion mining")	"all fields"

taxonomy (mutually exclusive, collectively exhaustive), objective ending conditions (8) and subjective ending conditions (5).

After the meta-characteristic (Step 1) and the ending conditions are determined (Step 2), the iteration process starts until all ending conditions are met. Building upon our domain knowledge gained by the deduced sample analysis, we use the conceptual-to-empirical approach for the first iteration. By reviewing the characteristics of existing applications of DL in NLP and existing domain knowledge from the CRISP-DM Methodology, we derived the following dimensions that seem explanatory for a taxonomy. We initially found and named the dimensions: *NLP Application*, *NN Architecture*, *Type of Embedding*, *Learning Technique* and *Performance Measures*. We found that the dimensions are similar to the CRISP-DM Methodology, an iterative process model which is technology neutral as well as industry independent [45]. Furthermore, it is an established standard for Data Mining (DM) and according to a poll conducted by the DM community KDNuggets, CRISP-DM is the most used model in

practice [46]. Originally the methodology includes of the following steps: (a) business understanding; (b) data understanding; (c) data preparation; (d) modeling; (e) evaluation; and (f) deployment [45, 27]. We renamed our initial dimension names of the dimensions to reflect the CRISP-DM Methodology. We would like to highlight that the term business understanding sets focus on understanding project goals and specifications from a business context, and then translating this information to a description of data mining problem definition [45]. We extend this understanding to also include a research-based application/understanding and rename it to "application understanding". This new term reflects on the development process of DNNs in NLP which do not necessarily have to include business aspects. We did not include deployment, because this term was insufficiently addressed in our sample papers and hence not relevant enough. Subsequently, this results in the following a preliminary taxonomy  $T_0$ .

**Table 3. Preliminary taxonomy**

$$T_0 = \left\{ \begin{array}{l} D_1 \text{ Application Understanding} = \{\text{Empty}\} \\ D_2 \text{ Data Preparation} = \{\text{Empty}\} \\ D_3 \text{ Modeling} = \{\text{Empty}\} \\ D_4 \text{ Learning Technique} = \{\text{Empty}\} \\ D_5 \text{ Evaluation} = \{\text{Empty}\} \end{array} \right\}$$

We identified several characteristics for each Dimension  $D_1 - D_5$ . In this first iteration, our goal was not to be concise but to include all possible characteristics for each dimension (i.e. collectively exhaustive but not mutually exclusive). The literature review yielded the following characteristics, which were then attributed to the corresponding dimensions  $D_1 - D_5$ . This step resulted in the first iteration of the taxonomy. All characteristics are granular aspects of the meta-characteristic defined in step 1.

The first iteration does not meet several ending conditions, but this was to be expected. Four dimensions were created in this iteration. Especially  $D_1$  does not meet various requirements such as mutual exclusivity or conciseness. In addition, only a fraction of the whole sample was included to create the first iteration. The taxonomy seems to be collectively exhaustive in terms of the characteristics in each dimension, but not for the dimensions. Although it fulfils the condition of having a limited number of dimensions it may be neither robust nor concise. This leads to the conclusion that at least one more iteration is needed.

We continued with another conceptual-to-empirical

iteration, followed by an empirical-to-conceptual and lastly a conceptual-to-empirical iteration. In this fourth iteration step, no new dimensions or characteristics were added. The entire sample of applied DL in NLP papers has been analysed. The taxonomy is concise, robust, comprehensive, expandable and explanatory. Hence, the taxonomy now meets the objective and subjective ending conditions. After all ending conditions are met, the development process is stopped. An overview of all iteration steps can be found in Table 4.

**Table 4. Ending conditions for iteration steps 1-4**

Iteration				Ending Conditions
1	2	3	4	<b>Taxonomy restriction definition</b>
<b>Definition of Taxonomy (2)</b>				
•	•	•		Mutually exclusive: No object has two different characteristics in a dimension
•	•	•		Collectively exhaustive: Each object has at least one characteristic in each dimension
<b>Objective Ending Conditions (8)</b>				
•	•			All objects (or a representative sample) were analysed
•	•	•		No object was merged or split in the last iteration
•	•	•		At least one object assigned to each characteristic
	•			No new dimensions or characteristics were added in the last iteration
	•			No dimensions or characteristics were merged or split in the last iteration
	•			Every characteristic within the dimension is unique
•	•	•		Every dimension is unique
•	•			Every combination of characteristics is unique
<b>Subjective Ending Conditions (5)</b>				
•	•			Concise: Dimensions and characteristics are limited
•	•	•		Robust: Sufficient number of dimensions and characteristics
•	•	•		Comprehensive: Identification of all (relevant) dimensions of an object
•	•	•		Extendable: Possibility to easily add dimensions and characteristics in the future
•	•	•		Explanatory: Dimensions and characteristics sufficiently explain the object

### 4.3. Final Taxonomy

After four development iterations, we arrive at the final taxonomy shown in Figure 3. The taxonomy contains five dimensions, each composed of several characteristics (Note that the characteristics' index is not consecutive because in the iteration steps some were dropped or combined into a new characteristic).

Since a taxonomy should not just be descriptive but explanatory [3], and to make interpreting the dimensions easier, we describe resulting the dimensions below:

- **D<sub>1</sub> Application Understanding:** Focuses

on understanding the project objectives and requirements and conversion into a NLP problem definition [45]. It lists "Current approaches to solving several ...useful problems in natural language processing" [20].

- **D<sub>2</sub> Data Preparation:** Covers the method used to represent discrete variables as continuous vectors. Tasks also may include tokenization, pruning or stopword removal.
- **D<sub>3</sub> Modeling:** Describes the design of the NN
- **D<sub>4</sub> Learning Technique:** Describes how a model improves at performing on some task with experience [47]
- **D<sub>5</sub> Evaluation:** Types of identification and quantification methods to evaluate a model's performance that enable a statement to be made about the extent to which objectives have been achieved in terms of quality, quantity or time

Correspondingly to the dimensions, we will elaborate on some characteristics which might not be self-explanatory. Especially  $D_1$  and  $D_3$  include more abstract namings of certain characteristics.

- **D<sub>1</sub> Application Understanding:**
  - C<sub>1,6</sub> Lexical Normalisation: The task of translating/transforming a non-standard text to a standard register
  - C<sub>1,30</sub> Semantic Annotation: Attaching additional information to various concepts in a given text
- **D<sub>2</sub> Data Preparation:**
  - C<sub>2,1</sub> CW: Collobert Weston
  - C<sub>2,8</sub> HLBL: Hierarchical log-bilinear
  - C<sub>2,9</sub> LSA: Latent Semantic Analysis
- **D<sub>3</sub> Modeling:**
  - C<sub>3,8</sub> DMN: Dynamic Memory Network

## 5. Discussion

To answer our research question: *What are the theoretically grounded and empirically validated key components of Deep Neural Networks used in Natural Language Processing?*, we iteratively derived a novel taxonomy for DL in NLP containing five dimensions with each between three and twelve characteristics. This taxonomy contributes to applied NLP by providing a systematically derived structure of NN characteristics

Dimensions $D_i$				
$D_1$ Application Understanding	$D_2$ Data Preparation	$D_3$ Modeling	$D_4$ Learning Technique	$D_5$ Evaluation
Characteristics $C_{ij}$				
$C_{1,6}$ Lexical Normalisation	$C_{2,1}$ CW	$C_{3,4}$ GRU	$C_{4,1}$ Supervised	$C_{5,1}$ Error Identification
$C_{1,7}$ Machine Translation	$C_{2,2}$ BERT	$C_{3,8}$ DMN	$C_{4,2}$ Unsupervised	$C_{5,2}$ Machine Translation Evaluation
$C_{1,10}$ Part-of-speech Tagging	$C_{2,3}$ dict2vec	$C_{3,10}$ GAN	$C_{4,3}$ DRL	$C_{5,3}$ Summarisation
$C_{1,11}$ Question Answering	$C_{2,4}$ doc2vec	$C_{3,11}$ (LSTM) RNN		$C_{5,4}$ Image Caption
$C_{1,20}$ Speech Recognition	$C_{2,5}$ ELMo	$C_{3,12}$ (LSTM) CNN		$C_{5,5}$ Parser Evaluation
$C_{1,27}$ Information Extraction	$C_{2,6}$ fasttext	$C_{3,13}$ Encoder/Decoder		$C_{5,6}$ Runtime
$C_{1,28}$ Corpus Analysis	$C_{2,7}$ GloVe			
$C_{1,29}$ Text Generation	$C_{2,8}$ HLBL			
$C_{1,30}$ Semantic Annotation	$C_{2,9}$ LSA			
$C_{1,31}$ Sentiment Analysis	$C_{2,10}$ ngram2vec			
$C_{1,32}$ Text Summarisation	$C_{2,11}$ word2vec			
$C_{1,33}$ Text Classification or Categorisation				

Figure 3. Final Taxonomy

in NLP which can be used to structure and organise knowledge in the field of DL in NLP.

To show the value of our taxonomy, we aim to display the application through an exemplary use-case following the method of Yin [48]. We decided to classify a recent paper by Borjali et al. [49] from the Information Systems (IS) domain to provide fellow IS researchers further guidance. We start with the first dimension  $D_1$  *Application Understanding*, where they try to detect medical adverse events from free-text medical narratives, this application falls into our characteristic  $C_{1,27}$  *Information Extraction*. They use  $C_{2,11}$  *word2vec* for embedding ( $D_3$  *Data Preparation*) and  $C_{3,11}$  *(LSTM) RNN* from the  $D_2$  *Modeling* dimension and. The authors labeled the radiology notes into three categories, thus using a  $C_{4,1}$  *Supervised* learning approach from dimension  $D_4$  *Learning Technique*. Lastly, the models were evaluated using quantitative measures such as precision and recall. They fall into our  $C_{5,1}$  *Error Identification* characteristic within the  $D_5$  *Evaluation* dimension. See Figure 4 for an overview.

Having shown an application of our taxonomy, it is important to note that researchers and practitioners might add new dimensions and/or characteristics. This is necessary because the field is dynamic. New upcoming architectures or evaluation techniques may lead to a violation of ending conditions (e.g. not collectively exhaustive), so the developed taxonomy may be extended to reflect such changes. This is a critical task to ensure the usefulness of any taxonomy. We aim to provide the investigated knowledge as an openly available web platform for fellow researchers and practitioners. Thus, we offer them an open-sourced set of techniques for them to design, compare and

evaluate existing NLP challenges and applications regarding DL.

In terms of limitations, a taxonomy’s endeavour to generality limits the granularity of both the dimensions and characteristics [3]. So the taxonomy might be sufficient for researchers and practitioners but not granular enough for specialised scholars. It is important to note, that the taxonomy has not been evaluated using interviews with researchers and practitioners from this field. What is more, we cannot exclude the possibility that, for a given model, multiple possible characteristics exist, which contradicts the definition of mutually exclusive characteristics [3]. This due to the “representative” nature of our coverage strategy of selectively covering works that were representative of the literature.

The presented taxonomy should be used while keeping in mind several assumptions and decisions made during its development. Especially the subjective selection criteria in the taxonomy development (selection of the relevant papers within the sample, as well as the selection and naming of dimensions and/or characteristics) may give rise to some bias. For other researchers and practitioners, other dimensions/characteristics may be important.

## 6. Conclusion

In this paper, we developed *A Taxonomy for Deep Learning in Natural Language Processing*. We conducted four consecutive development iterations according to the method presented by Nickerson et al. [3]. The taxonomy is grounded on a set of 205 articles on DL and NLP that we identified through a

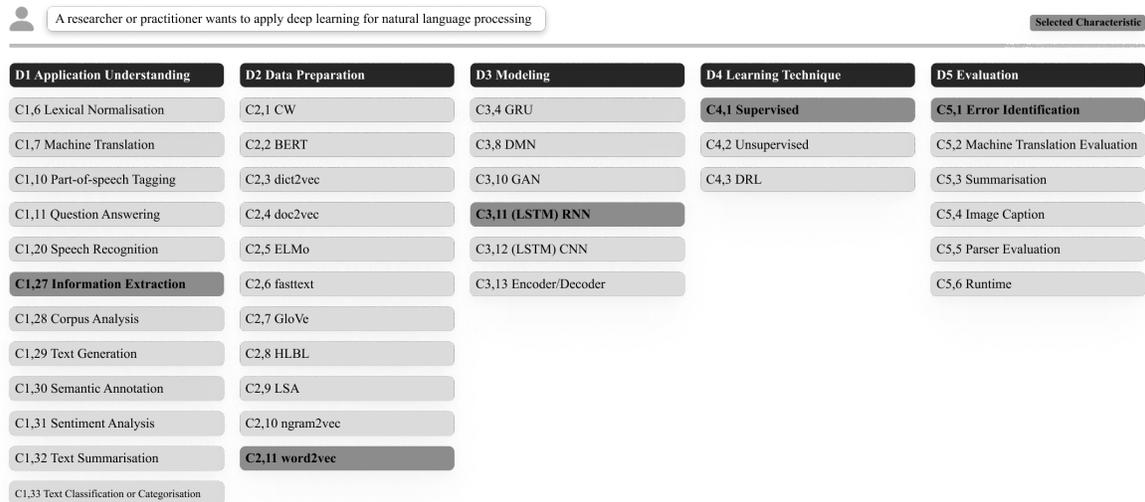


Figure 4. Applied Taxonomy

systematic literature review. We extend the literature on DL and NLP, by developing a new holistic taxonomy that structures and classifies key components of DNN from NLP perspective and thus goes beyond existing frameworks that focus primarily on a single dimension or stark technical focus.

For further research, we suggest changing the literature review's focus to also include more sources from the field of computational linguistics. The problem within applied NLP sources is that they are rather results-focused. Hence, the underlying models and architectures are not explained in sufficient detail. Whereas in computational linguistics papers, the used architectures are made clear more prominently. Additionally, the taxonomy might be used as a starting point for future research focusing on highly specialised taxonomies and additionally can be used to find and define archetypes of DNN for NLP. As discussed, the taxonomy's relevance for researchers and practitioners could be increased by conducting expert interviews to evaluate the findings.

Overall, we hope this taxonomy will be a valuable contribution to the general understanding and use of DL in NLP and support the further development of NLP research and application.

## References

- [1] J. Webster and R. T. Watson, "Analyzing the past to prepare for the future: Writing a literature review," *MIS Quarterly*, vol. 26, no. 2, pp. xiii–xxiii, 2002.
- [2] J. vom Brocke, A. Simons, B. Niehaves, K. Riemer, R. Plattfaut, and A. Cleven, "Reconstructing the giant: On the importance of rigour in documenting the literature search process," in *Information systems in a globalising world : challenges, ethics and practices ; ECIS 2009, 17th European Conference on Information Systems* (S. Newell, E. Whitley, N. Pouloudi, J. Wareham, and L. Mathiassen, eds.), pp. 2206–2217, Università di Verona, 2009.
- [3] R. C. Nickerson, U. Varshney, and J. Muntermann, "A method for taxonomy development and its application in information systems," *European Journal of Information Systems*, vol. 22, no. 3, pp. 336–359, 2013.
- [4] T. Bruecken, "Wie ein Startup aus Koeln Google Translate abhängen will," 2017.
- [5] Y. Bengio, "Learning deep architectures for ai," *Foundations and Trends in Machine Learning*, vol. 2, no. 1, pp. 1–127, 2009.
- [6] D. Graupe, *Principles of Artificial Neural Networks*, vol. 3 of *Advanced Series in Circuits and Systems*. World Scientific, 1997.
- [7] N. Buduma and N. Locascio, *Fundamentals of Deep Learning : Designing Next-Generation Machine Intelligence Algorithms.*, vol. First edition. O'Reilly Media, 2017.
- [8] T. Wambsganss, C. Niklaus, M. Cetto, M. Söllner, S. Handschuh, and J. M. Leimeister, "AI: An adaptive learning support system for argumentation skills," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, (New York, NY, USA), p. 1–14, Association for Computing Machinery, 2020.
- [9] T. Wambsganss, M. Söllner, and J. M. Leimeister, "Design and Evaluation of an Adaptive Dialog-Based Tutoring System for Argumentation Skills," in *International Conference on Information Systems (ICIS)*, (Hyderabad, India), 2020.
- [10] S. Ruder, "sebastianruder/NLP-progress," 2020. original-date: 2018-06-22T17:43:55Z.
- [11] J. Feine, U. Gnewuch, S. Morana, and A. Maedche, "A taxonomy of social cues for conversational agents," *International Journal of Human-Computer Studies*, vol. 132, pp. 138–161, 07 2019.
- [12] N. Zierau, T. Wambsganss, A. Janson, S. Schöbel, and J. M. Leimeister, "The anatomy of user experience with

- conversational agents: A taxonomy and propositions of service clues,” in *International Conference on Information Systems (ICIS)*, 2020.
- [13] H. Fromm, T. Wambsganss, and M. Söllner, “Towards a taxonomy of text mining features,” in *Twenty-Seventh European Conference on Information Systems (ECIS2019)*, June 2019.
- [14] “CS231n convolutional neural networks for visual recognition,” 2019.
- [15] D. Jurafsky and J. H. Martin, “Speech and Language Processing,” unpub.
- [16] Y. Goldberg, “Neural network methods for natural language processing,” *Synthesis Lectures on Human Language Technologies*, vol. 10, no. 1, pp. 1–309, 2017.
- [17] J. Eisenstein, “Natural Language Processing,” 2018.
- [18] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.
- [19] S. Zhang, L. Yao, A. Sun, and Y. Tay, “Deep learning based recommender system: A survey and new perspectives,” *ACM Comput. Surv.*, vol. 52, no. 1, pp. 5:1–5:38, 2019.
- [20] D. W. Otter, J. R. Medina, and J. K. Kalita, “A survey of the usages of deep learning in natural language processing,” *arXiv:1807.10854 [cs]*, 2018.
- [21] J. Thanaki, *Python Natural Language Processing*. Packt Publishing Ltd, 2017. Google-Books-ID: IedDDwAAQBAJ.
- [22] N. Andrews and E. Fox, “Recent developments in document clustering,” 11 2007.
- [23] K. Benoit, K. Watanabe, W. Haiyan, P. Nulty, A. Obeng, S. Mueller, A. Matsuo, L. Jiong-Wei, P. Perry, J. Kuha, B. Lauderdale, W. Lowe, C. Mueller, L. Young, S. Soraka, and I. Fellows, *quanteda: Quantitative Analysis of Textual Data*, 2018. R package version 0.7.6.
- [24] T. Mikolov, K. Chen, G. Corrado, and J. Dean, “Efficient Estimation of Word Representations in Vector Space,” 2013.
- [25] J. Pennington, R. Socher, and C. Manning, “Glove: Global vectors for word representation,” in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1532–1543, Association for Computational Linguistics, 2014.
- [26] “Arxiv submission rate statistics,” 2019.
- [27] T. Wambsganss, N. Molyndris, and M. Söllner, “Unlocking transfer learning in argumentation mining: A domain-independent modelling approach,” in *15th International Conference on Wirtschaftsinformatik (WI)*, pp. 341–356, 2020.
- [28] F. Möller, M. Stachon, C. Hoffmann, H. Bauhaus, and B. Otto, “Data-driven business models in logistics: A taxonomy of optimization and visibility services,” *Hawaii International Conference on System Sciences 2020 (HICSS-53)*, 2020.
- [29] M. Reuter-Oppermann and C. Wolff, “Towards a unified understanding of data-driven support for emergency medical service logistics,” *Hawaii International Conference on System Sciences 2020 (HICSS-53)*, 2020.
- [30] F. Kolini and L. Janczewski, “Cyber defense capability model: A foundation taxonomy,” *CONF-IRM 2015 Proceedings*, 2015.
- [31] M. Nusir and D. Bell, “Systematic literature review: Taxonomy of services in e-government,” *UK Academy for Information Systems Conference Proceedings 2013*, 2013.
- [32] M. Gardner and S. Dorling, “Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences,” *Atmospheric Environment*, vol. 32, no. 14, pp. 2627 – 2636, 1998.
- [33] S. Gollapudi, *Practical Machine Learning*. Community Experience Distilled, Packt Publishing, 2016.
- [34] C. E. journal, “Taxonomy of methods for deep meta learning – intuition machine,” 2017.
- [35] S. Ruder, “Neural transfer learning for natural language processing (PhD thesis),” 2019. Library Catalog: ruder.io.
- [36] S. Srinivas, R. K. Sarvadevabhatla, K. R. Mopuri, N. Prabhu, S. S. S. Kruthiventi, and R. V. Babu, “A taxonomy of deep convolutional neural nets for computer vision,” *Frontiers in Robotics and AI*, vol. 2, p. 36, 2016.
- [37] E. Aljalbout, V. Golkov, Y. Siddiqui, and D. Cremers, “Clustering with deep learning: Taxonomy and new methods,” *ArXiv*, vol. abs/1801.07648, 2018.
- [38] A. G. Blaiech, K. Ben Khalifa, C. Valderrama, M. A. C. Fernandes, and M. H. Bedoui, “A survey and taxonomy of FPGA-based deep learning accelerators,” *Journal of Systems Architecture*, vol. 98, pp. 331–345, 2019.
- [39] S. Ferilli, N. Di Mauro, T. M. A. Basile, and F. Esposito, *A Taxonomic Generalization Technique for Natural Language Processing*. Lecture Notes in Computer Science, Springer Berlin Heidelberg, 2011.
- [40] S. Dobnik and J. D. Kelleher, “Modular mechanistic networks: On bridging mechanistic and phenomenological models with deep neural networks in natural language processing,” *CoRR*, vol. abs/1807.09844, 2018.
- [41] J. Muntermann, R. C. Nickerson, and U. Varshney, “Towards the development of a taxonomic theory,” in *AMCIS*, 2015.
- [42] K. D. Bailey, “Typologies and taxonomies: An introduction to classification techniques,” *Journal of the American Society for Information Science*, vol. 47, no. 4, pp. 328–329, 1996.
- [43] H. M. Cooper, “Organizing knowledge syntheses: A taxonomy of literature reviews,” *Knowledge in Society*, vol. 1, no. 1, p. 104, 1988.
- [44] G. Al-Bdour, R. Al-Qurran, M. Al-Ayyoub, and A. Shatnawi, “A detailed comparative study of open source deep learning frameworks,” *arXiv:1903.00102 [cs]*, 2019.
- [45] “IBM SPSS modeler CRISP-DM guide,” 2016.
- [46] G. Piatetsky, “CRISP-DM, still the top methodology for analytics, data mining, or data science projects,” 2014.
- [47] J. Furnkranz, “Data Mining und Maschinelles Lernen Introduction,” 2017.
- [48] R. K. Yin, “Validity and generalization in future case study evaluations,” *Evaluation*, vol. 19, no. 3, pp. 321–332, 2013.
- [49] A. Borjali, M. Magneli, D. Shin, H. Malchau, O. K. Muratoglu, and K. M. Varadarajan, “Natural language processing with deep learning for medical adverse event detection from free-text medical narratives: A case study of detecting total hip replacement dislocation,” 2020.