A Review on Linguistic Style Matching

TOM POSTMUS, University of Twente, The Netherlands

Linguistic Alignment (LA) and Linguistic Style Matching (LSM) are phenomena in dialogue that have both been measured in many human dialogue studies. LA has also shown to be useful in Human Agent interaction. On the surface, LA and LSM seem similar, in the sense that they share conceptual characteristics. However, a detailed review of the similarities of LA and LSM is missing, as well as a contemporary review on the state of the art of LSM. In the current research, a Systematic Literature Review is executed to review the state of the art of LSM and similarities between LSM and LA. Furthermore, it is reviewed how LSM could be useful in the context of Human Agent interaction, similar to how LA sees its use in Human Agent interaction.

Additional Key Words and Phrases: Linguistic Style Matching, LSM, Linguistic Alignment, Human Agent interaction, conversational agent

1 INTRODUCTION

In the NL4XAI project[1], Linguistic Alignment is researched for its use in conversational agents to give explanations to the user. The current research is under supervision of the NL4XAI project and is interested in Linguistic Style Matching and its relation to Linguistic Alignment.

Linguistic Alignment (LA), also called linguistic entrainment [1], is the phenomenon in human dialogue that Dialogue Participants (DPs) tend to use the same words, pronunciations or sentence structures. Pickering & Garrod in 2004 [2] gave LA its name, based on the observation that DPs *align* their word use, pronunciations and sentence structures over time.

LA was shown to occur in human dialogue, from now referred to as Human Human dialogue, or HH dialogue, by many studies [3, 4, 5]. In these studies LA was quantified using various computational measures [6]. These measures rely on the counting of similar words, or word groups, that are used between DPs.

LA was shown to occur also in the context of Human Agent (HA) interaction [7]. LA in HA interaction has various positive implications such as increased user engagement [8] and increased variation of the dialogue [9]. The following examples of LA in HA interaction are known:

1. Spillner & Wenig 2021 [8] implemented an alignment effect in an agent and showed that their agent increased the engagement of the user and decreased the perceived workload for the task at hand.
2. Dušek & Jurčíček 2016 [10] implemented a neural network that generated language that was aligned to the user. They showed that users preferred their aligning agent.
3. Levitan et al. 2016 [11] implemented an agent that aligned to the prosodic features (i.e. acoustic features) of the user's speech. They showed their agent was perceived as more likable and reliable by users.

Linguistic Style Matching (LSM) is another phenomenon in dialogue, named by Niederhoffer & Pennebaker in 2002 [12]. LSM is the similarity of language styles between DPs, i.e. DPs match their word use. Similar to LA, LSM has been quantified by computational measures in many studies on HH dialogue [13, 14, 15].

LSM and LA are in fact very similar concepts [16]. They share some conceptual characteristics:

1. LSM and LA are both phenomena of shared language (see Section 3.2).
2. LSM and LA happen both subconsciously in people [17, 2].
3. LSM and LA both have positive outcomes for the dialogue, such as increased rapport between DPs [16].

However, LSM and LA have not yet been compared to each other in a detailed review. Furthermore, it is the author's impression that LSM could also be implemented in a conversational agent; similar to how LA is implemented in agents. However, research of the implementation of LSM in conversational agents is rare.

The current research is aimed to bridge these gaps. First, a review on the state of the art of LSM is conducted, to get a clear picture of its applications and workings. To this end, it is first reviewed how it was initially defined. Then it is reviewed where LSM has been applied since its definition. Here it is also explored what techniques are used in the measurement of LSM. After having a clear picture of the state of the art of LSM, it is compared to LA and it is reviewed how LSM could be of use in the context of HA interaction.

The research is aimed at answering the following research questions:

RQ1 *What is the state of the art of LSM?*
- **[a]** *How has LSM been defined?*
- **[b]** *What has LSM been used for since its definition?*
- **[c]** *How is LSM measured?*

RQ2 *How does LSM relate to Linguistic Alignment?*
RQ3 *How could LSM be useful in the context of HA interaction?*

The current research follows a Systematic Literature Review (SLR) procedure, inspired by Pati et al. 2018 [18]. In this procedure, literature is searched and reviewed in a systematic way. This way, a broad selection of papers is considered. Furthermore, the systematic nature of the research makes it easier to document the steps that for taken. This in turn makes it easier to be evaluated, e.g. by peerreviewers.

The the next section (Section 2), the SLR procedure that was used in this research is explained. It is explained how the literature was searched and how it was then interpreted in order to answer the research questions. In the section after that (Section 3), the results of the SLR are stated. This includes the quantities of papers that were found and (most importantly) the interpretations that were

[1]NL4XAI project: https://nl4xai.eu/

38th Twente Student Conference on IT, February 3rd, 2023, Enschede, the Netherlands

derived from the papers that were found. Then follows a discussion section (Section 4), where is briefly reflected on the key arguments made in this article. Lastly, there is the conclusion section (Section 5) that summarizes the research.

2 METHOD

As the first step in the current SLR, the definitions for the search phase were defined. These definitions include: which databases to search, the relevant concepts to search for, the search keywords to use and which inclusion/exclusion criteria to follow. In the following section an overview is given of the search definitions.

2.1 Search definitions

There were three databases chosen for this SLR:

1. Web of Science[2]: this is a large and general database. Suitable for searching across multiple domains for papers relevant to LSM or LA.
2. ACL Anthology[3]: this is a more niche database, specific to computational linguistics. Convenient for finding papers on LSM or LA in the context of HA interaction.
3. Google Scholar[4]: this is a general database suitable for broad searches. Google Scholar also shows non-peerreviewed materials and other types of materials such as books or talk introductions.

Then, the following concepts were identified as relevant to search for based on the research questions:

1. Linguistic Style Matching.
2. Linguistic Alignment.
3. Entrainment: a concept close to alignment and they are used interchangeably.
4. Coordination: two or more people converging to the same behaviour.
5. Adaptive agent: an agent that adapts to the user.

Based on these relevant concepts, search queries were created to be used in the databases. Each database works differently and thus the queries are tailored to each database separately to give managable results [18]. See table 2 in the Appendix for a complete list of the search queries that were used and how many results were found.

2.2 Search procedure

The procedure of the search phase of the SLR was as follows: when a search query was entered, the list of results was screened, top to bottom, based on their titles. If the title was an indication that the paper was about a different concept, the paper would not be included. If the title seemed to point to a relevant topic, the abstract of the paper was read. Then, the paper would again be judged for inclusion or exclusion. This procedure is a slight variation on Pati et al. [18], since it merges the title screening phase and abstract screening phase into one phase, i.e. the results are iterated though once instead of twice. The decision to merge these phases was made because of time restrictions on the current research.

[2]Web of Science: https://clarivate.com/webofsciencegroup/solutions/web-of-science/
[3]ACL Anthology: https://aclanthology.org/
[4]Google Scholar: https://scholar.google.com/

2.2.1 End conditions of search. For each search query, the list of results was screened from top to bottom, until no relevant sources were found anymore. That is, until there were around five to ten irrelevant papers since the last relevant paper. The search phase would be terminated until it was determined that all relevant concepts were sufficiently explored, i.e. all potentially relevant search queries were used. In reality, the search was cut a bit short due to time restrictions. Fortunately however, there was already a substantial collection of relevant papers.

2.2.2 Structuring of Journal Citation Report. The papers that were deemed relevant, based on both their title and abstract, were included in the Journal Citation Report (JCR) [18]. More specifically, the papers were deemed relevant if they belonged in one of the following categories:

1: The paper is about LA or LSM, however is not in the context of HA interaction, but in the context of HH interaction.
2: The paper is about a different linguistic concept than LA or LSM, but in the context of HA interaction.
3: The paper is about LA or LSM. Furthermore, the paper is about the application of such a concept in HA interaction, but it has not implemented an agent.
4: The paper is about LA or LSM. Furthermore, the paper is about the implementation of such a concept in an agent.
5: The paper talks about theoretical notions of LSM.

The papers were given a *category index* (corresponding to the numbers above) in the JCR, so that the JCR would be structured and more easily navigable.

The above mentioned categories were chosen based on the initial intuition of the author, with some refinements made during the search process. The motivations of the way the categories chosen, were as follows: (1) The first category was chosen since it became apparent that papers on the application of LSM or LA on HH dialogue, were very numerous. Thus these papers needed to be separated, as to not ‘engulf‘ the smaller amount papers on LSM or LA in HA interaction. Furthermore, this category was also chosen for its relevance to assess the state of the art of LSM, i.e. research question **RQ1**, and a comparison between LSM and LA, i.e. research question **RQ2**. (2) Category two was formulated since it is potentially interesting to see a technique applied to HA interaction, but from the perspective of another theoretical framework, such as Communication Accommodation Theory or behavioural mimicry, which are similar to LA and LSM [16]. This could inspire creative views on the implementation of LSM in HA interaction, thus relevant to research question **RQ3**. (3) (4) The third and fourth category are both about the application of LSM and LA in HA interaction. Thus these papers are relevant for answering research question **RQ3**. These were made distinct categories however, since it was deemed more relevant to **RQ3** if the paper discussed an implementation of LSM or LA in an agent, i.e. category four. This category would later be read in detail (see next section) while category three would not, due to time restrictions and a lesser relevance to the research. (5) The fifth category was chosen, since papers that directly talk about theoretical notions of LSM are a good source for understanding its workings and its state of the art. This includes papers that define,

improve, clarify or disprove notions of LSM. Hence, these papers are relevant to all research questions.
It is important to note that not all categories are mutually exclusive. Namely, there is overlap between category one and five. For example, it is possible that a paper makes conclusions on the nature of LSM and also conducts a study of LSM in HH dialogue (e.g. Niederhoffer & Pennebaker 2002 [12]).

2.3 Interpretation of papers

After the search phase, the collection of papers in the JCR was (partly) summarized and interpreted. The papers that were given a category index one, two and three were summarized only based on their abstracts, because of their large numbers. Additionally, from category one papers it was also checked what methods of measuring LSM or LA they use in practice, which is relevant for exploring the state of the art of LSM (**RQ1**) and a practical comparison between LA and LSM (**RQ2**). Papers from categories four and five were read through in detail and summarized. These categories received more attention since they were chosen such that they are most relevant to the research questions. The summaries and other data (e.g. the LSM methods used) were then interpreted in order to answer the research questions.
In the following section these interpretations are covered.

3 RESULTS

The search phase of the SLR yielded a total amount of 146 papers. Among the categories (see Section 2.2.2) category one contained the most papers, namely 104. This can be explained by the large amount of studies that researched LSM and LA in the context of HH interaction. Category two contained 18 papers and category three contained 14. Category four contains 8 papers. Category five contains 6 papers. Recall that there is some overlap between category one and five (namely 4 papers). Hence the individual categories add to more than the total amount of papers.
From category four, only 2 papers discuss the implementation of LSM in HA interaction, namely Hoegen et al. 2019 [19] and Thomas et al. 2020 [20]. This suggests that LSM in conversational agents is indeed a little researched application.

In the following sections, the research questions are answered on the basis of the papers that were found. There are cases were papers are cited that were originally not part of the JCR (i.e. not found in the search phase), but that were retrieved from references of papers from the JCR. This process of finding papers on the basis of references of other papers is called *snowballing* [21].

3.1 The State of the Art of LSM

3.1.1 How has LSM been defined? LSM was first mentioned by Niederhoffer & Pennebaker in 2002 [12]. In their work, they defined the phenomenon of LSM as follows:

Definition 3.1. The words of one DP influence the other DP to respond in a certain way on a turn-based level and the broader conversational level.

This is a rather loose definition since it does not specify *how* the words of one DP influence the other DP. Or what is meant by the broader conversational level. Presumably, the definition was unspecific since Niederhoffer & Pennebaker 2002 [12] did not yet have concrete evidence of DPs matching their language style in dialogue. Hence, they conducted an experiment to find out more about LSM in dialogue. In the experiment, they calculated the similarity of word use between participants of text chat conversations [12]. Central to this calculation method was the The Linguistic Inquiry Word Count (LIWC, pronounced 'luke') software developed by Pennebaker, Booth & Francis [22]. The LIWC program analyses a text based on different word categories. For each category it calculates the percentage of words it takes up in the total amount of words in the text. The categories include both linguistic categories (pronouns, articles) and psychological categories (emotions, thinking styles). The categories were developed with the aid of judges and their validity was confirmed by over a 100 studies [23].
Niederhoffer & Pennebaker 2002 [12] used LIWC to count words for each DP. Then they correlated the categories that had the best reliability over time between the DP, i.e. that remained the most constant within and between conversations. For each category they calculated the correlation between the DPs (correlation coefficient). These correlation values would be LSM scores or LSM values between two DPs.
Niederhoffer & Pennebaker 2002 [12] calculated the LSM scores in their experiment in order to test the hypothesis that higher LSM is related with increased rapport. Although they did find high values of LSM, i.e. similarity of language styles, it was not correlated with perceived increased rapport by the participants.

3.1.2 What has LSM been used for since its definition? Since Niederhoffer & Pennebaker in 2002 [12], LSM has been applied as a methodology in many different settings. More specifically, similar procedures to Niederhoffer & Pennebaker 2002 [12] have been applied to measure LSM scores in HH dialogues. In the subsequent research to Niederhoffer & Pennebaker 2002 [12] the definition of LSM has also changed slightly. In subsequent research, researchers commonly refer to LSM as defined as follows:

Definition 3.2. LSM is the matching of *function words* between DPs [24, 25, 23, 26].

For the remainder of this article, we will consider LSM as defined by this definition. In the definition, function words are words that by themselves carry little meaning but serve a grammatical or structural role in a sentence [27]. Examples of function words are *pronouns, prepositions and articles.*
Psychologically, function words are uttered more subconsciously and automatically [25] and are tied to the psychological state of a person. For this reason, function words are also called *style words* since they are a good marker of the conversational style of a person [22].
Function words are the opposite of *content words*, which are words that carry more information. For example, they refer to an object or a concept [27]. Since LSM ignores content words, it more about *how* things are said, rather than *what* is said [16]. The new definition is based on the methodology that Niederhoffer & Pennebaker used in 2002 [12]. This methodology was based on correlating certain LIWC dimensions between DPs (see Section 3.1.1). However, these

LIWC dimensions were not specific to function words. Thus in this way, the definition is slightly different. The change in definition was likely motivated by Chung & Pennebaker in 2007 [17], where they connected LSM research with the concept of function word usage.

LSM is mostly applied as a methodology by researchers. That is, most researchers use LSM to measure the degree of style matching between DPs. Indeed, there are theoretical ideas that lay at the foundation of LSM. For example, that LSM between DPs reflects that they are in harmony with their psychological worlds [12]. Or that the use of function words is tied to psychological processes [17]. However, these theoretical ideas do not appear in the 'focus' of LSM research: most researchers use LSM as a methodology. The results of this SLR reflect that: 51 papers mentioned LSM. 47 of these papers used LSM as a methodology, which is around 92%.

Examples of settings where LSM was applied are:

1. LSM was measured between users of online forums [28, 24, 29, 30].
2. LSM was measured between client and therapist in therapy sessions [31, 32, 33, 13].
3. LSM was measured between politicians in debates or diplomatic negotiations [12, 14, 15]

Generally in LSM studies, LSM is measured on conversations of people and correlated with a social or psychological effect. For example, high LSM values in social group discussion is related to increased cohesiveness of this group [34]. Or being reminded of their mortality is correlated with increases LSM in conversations between people [35]. As mentioned, LSM is also applied on online forums. For example, high LSM between an individual coping with illness and other users of an online forum, contributes to an increased sense of perceived social support of the individual [29]. Or high LSM between users of an online community is related to a sense of identity with that community [24].

3.1.3 How is LSM measured? Methodologies of LSM that are used by researchers, are similar to each other. In fact, most of the methodologies build on the procedure used by Niederhoffer & Pennebaker in 2002 [12] (see Section 3.1.1). Gonzales et al. in 2010 [34] proposed a refined measure for LSM, that was based on the measure from Niederhoffer & Pennebaker 2002 [12] . This measure was also applicable on group interactions, i.e. LSM could be measured in conversation of more than two participants. Subsequently, Ireland & Pennebaker 2010 [36], made a slightly adapted version of the measure from Gonzales et al.
These two measures are used the most in subsequent LSM research for calculation of LSM: in the current research, among all articles that were found where LSM was measured in HH interaction, 36% of used the measure from Gonzales et al. and 55% from Pennebaker & Ireland.

Among the other 11%, there is Danescu-Niculescu-Mizil et al. 2011 that developed a probabilistic formula to measure LSM in a big-data setting such as Twitter [37]. In this formula it is taken into account how adjacent comments, i.e. who replies to whom, relate to each other stylistically. The fact that language style is correlated on a turn-by-turn basis, implies that a temporal component of LSM is measured [16]. Even though this method was developed for an online environment (i.e. Twitter), with comment feeds, it is also used in human conversation [31].

In 2018 Müller-Frommeyer et al. 2019 [16] reviewed the above measures of LSM and rated them on the basis of a few qualities. They pointed out that the measure from Pennebaker & Ireland (and thereby also the measure from Gonzales et al., due to similarity) lacked the consideration of the temporal component. These measures considered the text as a static whole, rather than considering the influence of utterances on a turn-by-turn basis. As mentioned, Danescu-Niculescu-Mizil et al. 2011 does implement this property in their LSM measure, however their measure lacks *frequency sensitivity*. That is, it does not account for the relative contributions of smaller LIWC categories. For example, for a smaller LIWC category, e.g. that takes up 4% of the text, a variation of 2% is more meaningful than for a category that takes up 20%. In addition to lacking frequency sensitivity, Müller-Frommeyer et al. 2019 [16] argued that the measure from Danescu-Niculescu-Mizil et al. 2011 was too complex in its usage due to the sophisticated nature of the probabilistic framework. Thus, it is not easily replicable.

Müller-Frommeyer et al. 2019 then proposed a measure rLSM, i.e. reciprocal LSM, that fulfils all of the desired properties. To the knowledge of the current research, the measure has since been employed in only two studies [32, 38].

In essence, all of the methodologies of measuring LSM are very similar. All of the above mentioned procedures, use LIWC to count the word categories. The counted LIWC categories are then fed into a formula that calculates the LSM score. This procedure is truly the backbone of LSM research.

3.2 LSM in relation to LA

LSM and LA are both phenomena of dialogue.

LSM: The phenomenon of LSM is defined as the matching of function word use between DPs (definition 3.2).

LA: The phenomenon of LA is defined as the alignment of language of DPs over the course of the conversation [2].

What can be immediately seen is that LSM and LA are both phenomena of shared language. Müller-Frommeyer et al. 2019 [16] argued the similarity of LSM and LA, specifically on the notions that they both happen subconsciously in people and both entail positive outcomes of dialogue, such as increased rapport.
In the next sections some differences are highlighted between LSM and LA.

3.2.1 LA is more general. A conceptual difference between LSM and LA, is that LA is a more general phenomenon. LSM is about the similarity of word use, while LA is about the similarity of language in general, i.e. words, sentence structures, pronunciations and concepts. It could be argued that LSM is akin to *lexical alignment*, which is the specific form of LA considering the alignment of words. However still, lexical alignment is different from LSM, since LSM is specifically about function words. Lexical alignment it is not restricted to only function words.
Put in other words, LSM is about *how* things are said, rather than *what* is said. LA is about both *how* things are said, as well as *what* is said.

3.2.2 Dynamic vs. static. Another conceptual difference is that LA is a *dynamic* phenomenon by definition, while LSM is not. Müller-Frommeyer et al. 2019 [16] clarify useful terminology with *similarity* versus *accommodation*. Similarity is a static notion of shared language. It refers to the overlap of language between DPs when considering the conversation as a whole, or piece of the conversation. Instead, accommodation refers to the *change* of similarity of language over time and is thus a dynamic concept. Accommodation is used interchangeably with the term *coordination*.

The dynamics in LA come from the notion that DPs align with their language over time. So alignment entails change in similarity of language between DPs over the course of the conversation. The fact that it is a phenomenon that develops over time makes it a dynamic phenomenon like accommodation. The definition of LSM does not regard any notion of time: it does not specify any development of function word use over time. Therefore, it is interpreted here as a static phenomenon. In terms of Müller-Frommeyer et al. 2019 [16]: it is a phenomenon of language similarity.

The most frequently used LSM measures from Gonzales et al. 2010 [34] and Ireland & Pennebaker 2010 [36] are indeed similarity measures, i.e. static measures. However, as discussed previously (Section 3.1.3) Müller-Frommeyer et al. 2019 [16] states that there are also LSM measures that measure a temporal component of style matching. That is, they measure the accommodation of language. Here it is argued it is better to call these measures as measures of Linguistic Style Accommodation (LSA). LSA can be seen as the dynamic cousin of LSM and is also an established term in research of linguistics [39, 40, 41].

3.2.3 Psychological models. The phenomena of LSM and LA both have theories on the psychological mechanics that are at play underneath the surface of the phenomena. Pickering & Garrod 2004 [2] defined the Interactive Alignment Model (IAM) that explains the psychological mechanics behind the phenomenon of LA. In the IAM, LA is caused by the convergence to equal *linguistic representations* of the DPs. A linguistic representation is a general term for a linguistic object, such as a word, a sentence, a pronunciation or a concept, as represented in the mind of a DP.

Central to the IAM, is the *priming mechanism*. This mechanism drives the alignment between DPs. It entails that for DP A, hearing an utterance from DP B is likely to activate the same linguistic representation in DP A. Thus this linguistic representation becomes aligned between DPs A and B. This in turn makes it more likely that the DP A will produce a similar utterance later in the conversation. This explains the phenomenon of LA that DPs come to produce similar utterances.

According to the IAM, alignment occurs at different levels. That is, DPs align on their use of words, sentence structures, pronunciations and concepts. This is due to the generality of the linguistic representation: it can be any type of linguistic representation (word, sentence structure, etc) that DPs align on.

For LSM such a model does not exist. There are psychological theories given that explain parts of the phenomenon. Chung & Pennebaker in 2007 [17] discuss theories about function words. They state that function words are uttered subconsciously and that they are correlated with social processes. However, these theories do not explain LSM to the same degree of completeness that the IAM does for LA. For example, Niederhoffer & Pennebaker 2002 [12] state that LSM is correlated with DPs being in harmony with their psychological worlds. However, it is not clear what psychological mechanics are the *cause* of LSM. Meanwhile, for LA it is the priming mechanism that causes alignment.

3.3 LSM in Human Agent Interaction

The conceptual similarities between LA and LSM suggest that LSM could also have its use in HA interaction. Thomas et al. in 2018 analysed interactions between human and agents and found that humans match their language style to that of an agent [42]. Indeed, this implies the possibility for LSM to occur in HA interaction, however with a missing piece of the puzzle: the agent should also the match language style of the user in order for there to be LSM in the conversation.

3.3.1 Current techniques. Hoegen et al. 2019 attempted the creation of such a style matching agent [19], i.e. an agent that exhibits LSM. Their agent could sense the user's language style and produce utterances that were consistent with this style. The agent was implemented as a chain of several components: the voice of the user was interpreted by a speech recognition module. A dialogue generator, that was trained on Twitter conversations, was used to generate *chit-chat* responses, i.e. responses that do not have much intent but give an impression of human-like small-talk. An intent recognizer was used to detect intents from the user input. These intent were related tot the task setting the dialogue was based in. For example, when the user is tasked to ask the agent for directions to the train station and they ask "What way is the train station?", the agent should recognize that. The responses that the agent would then give to each intent was scripted.

The language style was extracted from the user input, and calculated in the form of the variables as defined by Thomas et al. 2018 [42]. At the speech synthesis component, i.e. the component that forms the output speech of the agent, these language style variables would also be applied to the agent's speech. This way the output of the agent would match the language style of the input. Thus the agent exhibits LSM.

Hoegen et al. 2019 conducted an experiment where one group of participants interacted with an agent that exhibited LSM and another group interacted with agent that did not exhibit LSM. In their experiment, Hoegen et al. 2019 found that certain users rated the agent to be more trustworthy. These were users with a High Consideration conversational style as defined by Tannen [43].

Thomas et al in 2020 [20] implemented an LSM exhibiting agent using an adaptation of the architecture of Hoegen et al. 2019 [19]. They found that the style matching behaviour of the agent made for smoother conversations.

To the current research, Hoegen et al. 2019 and Thomas et al. 2020 are the only known implementations of LSM in HA interaction. They found meaningful results, however they acknowledge some shortcomings of their implementations concerning the naturalness of the dialogue. Of course it is not an easy task, due to the novelty of the application of LSM in HA dialogue [19, 20].

In the remainder of this section, two other possible approaches are proposed for the implementation of LSM in a conversation agent. The approaches are sketched on a high-level.

3.3.2 Personality model. Another way that LSM could be implemented in a conversational agent, could be through the use of a personality model. Mairesse & Walker in 2010 implemented a model for generation of stylized text based on a personality model [44]. This personality model was a set of values also known as the *Big Five traits*. This is a model in psychology where a personality is characterised by five traits: *extraversion, emotional stability, agreeableness, conscientiousness and openness to experience* [45]. Based on of the personality model, Mairesse & Walker defined a mapping to conversational style parameters. That is, certain values of the Big Five traits, would map to certain conversational style parameters. For example, high extraversion maps to frequent use of stop words. The motivations for the mappings were backed by various psychological research works [44].
The language style parameters would then be used in the generation of styled language.
The implementation of Mairesse & Walker is not suitable for LSM in HA interaction, since it only deals with *generation* and no *detection* of personality. That is, the personality traits are not detected automatically in real-time, but have to be manually fed in. Hence the language style of the agent would not adapt automatically to the user.
However according to Tausczik & Pennebaker in 2010 [22], some LIWC categories correspond directly to Big Five traits. Thus, LIWC could potentially be used to detect Big Five traits from the user's speech, which could in turn be fed into the stylized language generation module by Mairesse & Walker. This way, a style matching agent could be implemented that matches the personality of the user. One caveat here is that LIWC has never been used in a real-time computational pipeline. So a real-time version or adaptation of the LIWC algorithm would be necessary for such an implementation.

3.3.3 Big data approach. Another way to implement LSM in a conversational agent, could be a big data approach. Dušek & Jurčíček 2016 implemented entrainment in an agent using a neural network model that was trained on a data set of utterances with preceding context utterance [10]. The utterances were related to the public transport system of Manhattan. The preceding utterance referred to the utterance that was produced right before the current utterance. The neural network would learn from each preceding utterance what the appropriate style of the response should be. The style of the utterances was analysed using the *BLEU* measure by Papineni et al. [46].
This measure is actually intended for evaluating the quality of machine translations, e.g. in their fluency or adequacy. It is not immediately clear how Dušek & Jurčíček 2016 used this measure to compare the style of the utterances. Perhaps they treated the utterance as an English to English translation of the preceding utterance. Then the quality of this 'translation' could correspond to the quality of the correspondence of style.
In any case, the method of training a neural network on a database with utterances and corresponding preceding utterances, could be a valid approach for implementing LSM in an agent. A measure inspired by the current methodologies of measuring LSM (see Section 3.1.2) could be used to train the neural network. Such a measure could be an adaptation of the LIWC algorithm with the corresponding LSM calculation measure. Then, the neural network could be trained to generate responses with high LSM levels with the user.

4 DISCUSSION

In this section, it is reflected on the current research and what could have gone better in the SLR. Then, the main arguments made in the current research are listed and briefly reflected upon.

4.1 The current SLR

The current research was inspired on the SLR procedure from Pati et al. [18]. An important advantage of the SLR procedure is the systematic nature of the research that is conducted. This ensures both reportability and replicability. Another key advantage would be a smaller likelihood for biases, due to the between-members verification steps of the SLR procedure. However, in the current research, these verification steps were not executed due to the limited time available.
The time restrictions on the current SLR also restricted the broadness of the search phase. During the search phase there were more queries with potentially interesting results that could not be tried. An example of such a query, is any query related to Linguistic Style Accommodation (LSA). It was discovered that LSA is the dynamic variant of LSM (see Section 3.2.2). Thus, it could have been interesting to look further into LSA, e.g. for answering **RQ3** from perspectives of LSA.

4.2 Arguments

4.2.1 Redefinition of LSM. Despite the definition of LSM by Niederhoffer & Pennebaker in 2002 [12], LSM was interpreted differently by researchers of later work. Thus LSM underwent a slight redefinition. Where the original definition was rather theoretical, the new definition was based on the practical application of LSM, i.e. the measurement of similarity of function words of DPs. Perhaps, it was the fact that Niederhoffer & Pennebaker in 2002 [12] in the same article defined a procedure for measuring LSM, that inspired subsequent research to treat LSM as a measurement procedure.

4.2.2 Little variance in LSM methods. It was found that the procedures used for measuring LSM, are very similar to each other. In fact, all of them follow the same structure: analysis of word use with LIWC, then applying a formula on the measured LIWC categories. What formula is used can differ among the methodologies, however in most cases the formula from Gonzales et al. 2010 [34] and Ireland & Pennebaker 2010 [36] is used. Since LIWC is used in all of the methodologies, it can be seen as a staple of LSM research.

4.2.3 LA is more general than LSM. When comparing the phenomena of LSM and LA, it became apparent that LA is a phenomenon that covers a more general range of human interaction. Where LSM entails sharing the same function words, LA entails sharing same language.

4.2.4 LSM is static, LA is dynamic and LSA is dynamic.

From analysing the way LA and LSM are defined, a difference became apparent in their dynamics. Namely, that LA is a dynamic phenomenon and LSM a static phenomenon. The fact that LA is dynamic is explained by the notion that LA is a phenomenon that develops over time. For LSM, there is no mention of time and thus the phenomenon is called static. If a measure measures the development LSM between DPs over time, it is better to call this a measure of LSA, which is the dynamic version of LSM.

4.2.5 LA is explained more completely.

LA and LSM both have theories about the psychological mechanics that are at play underneath the surface. In the case of LA, there is the IAM model that explains LA in a complete way. That is, from the model it is clear why LA happens and, for example, why it happens at different levels.
For LSM, there are theories that explain certain facets of the phenomenon, such as that function words are uttered largely automatically. However, the set of theories on LSM do leave more gaps than the IAM does for LA. For example, it is not clear what psychological mechanism causes LSM.

4.2.6 LSM is useful in HA interaction.

There is some evidence that suggests that the application of LSM in conversational agents could be useful. Namely Thomas et al. 2020 [20] found that an LSM exhibiting agent makes for smoother conversations. Hoegen et al. 2019 [19] found that certain uses rated an LSM exhibiting agent to be more trustworthy.
Despite the fact that the evidence is not numerous, there is reason to believe that the application of LSM in a conversational agent could useful due to the similarities of LSM an LA: the positive outcomes that an LA exhibiting agent has, could also exist for an LSM exhibiting agent. For example, the fact that interacting with an LA exhibiting agent increases the engagement of the user could signify that an LSM exhibiting agent also increases the user engagement, since both LSM and LA entail that the language is shared between the DPs.

4.3 Future work

In section 3.3, two approaches were sketched on a high-level that could be implemented in order to achieve an agent that exhibits LSM. In future work, these approaches could be further concretised and implemented. Furthermore, future work could investigate more about the research on LSA and whether it has prospects in the implementation in conversational agents.

5 CONCLUSION

In this research, a review was conducted on the state of the art of LSM research. It was investigated how LSM was defined and how it was applied. The state of the art of LSM was explored in order to compare LSM to LA: two similar phenomena of human dialogue. Furthermore, it was explored what uses LSM could have in the implementation of a conversational agent, similar to how LA is employed to make HA dialogue more variable and engaging. The review was conducted in a systematic way: the search was executed systematically and the resulting papers were interpreted systematically. Important findings form the SLR were that LSM research is mostly focused on the methodological application of measuring LSM and that the method of measuring LSM is almost always the same. Furthermore, some conceptual differences of LSM and LA were found: LA is a more general phenomenon than LSM and LA is dynamic and LSM static. It was also explored what uses LSM could have in the context of an implementation in HA dialogue. Two existing methods were reviewed and two possible approaches for future research were sketched.

REFERENCES

[1] Andreas Weise, Vered Silber-Varod, Anat Lerner, Julia Hirschberg, and Rivka Levitan. 2021. "talk to me with left, right, and angles": lexical entrainment in spoken hebrew dialogue. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics.*

[2] Martin J Pickering and Simon Garrod. 2004. Toward a mechanistic psychology of dialogue. *Behavioral and brain sciences*, 27, 2, 169–190.

[3] Thomas Misiek, Benoit Favre, and Abdellah Fourtassi. 2020. Development of multi-level linguistic alignment in child-adult conversations. In *Proceedings of the workshop on cognitive modeling and computational linguistics*, 54–58.

[4] Gabriel Doyle and Michael C Frank. 2016. Investigating the sources of linguistic alignment in conversation. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 526–536.

[5] Yang Xu, Jeremy Cole, and David Reitter. 2018. Not that much power: linguistic alignment is influenced more by low-level linguistic features rather than social power. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 601–610.

[6] Yang Xu and David Reitter. 2015. An evaluation and comparison of linguistic alignment measures. In *Proceedings of the 6th workshop on cognitive modeling and computational linguistics*, 58–67.

[7] Holly Branigan and Jamie Pearson. 2006. Alignment in human-computer interaction. *How people talk to computers, robots, and other artificial communication partners*, 140–156.

[8] Laura Spillner and Nina Wenig. 2021. Talk to me on my level–linguistic alignment for chatbots. In *Proceedings of the 23rd International Conference on Mobile Human-Computer Interaction*, 1–12.

[9] Guillaume Dubuisson Duplessis, Caroline Langlet, Chloé Clavel, and Frédéric Landragin. 2021. Towards alignment strategies in human-agent interactions based on measures of lexical repetitions. *Language Resources and Evaluation*, 55, 2, 353–388.

[10] Ondřej Dušek and Filip Jurčíček. 2016. A context-aware natural language generator for dialogue systems. *arXiv preprint arXiv:1608.07076.*

[11] Rivka Levitan, Stefan Benus, Ramiro H Gálvez, Agustın Gravano, Florencia Savoretti, Marian Trnka, Andreas Weise, and Julia Hirschberg. 2016. Implementing acoustic-prosodic entrainment in a conversational avatar. In *Interspeech*. Vol. 16. San Francisco, CA, 1166–1170.

[12] Kate G Niederhoffer and James W Pennebaker. 2002. Linguistic style matching in social interaction. *Journal of Language and Social Psychology*, 21, 4, 337–360.

[13] Rachel Schaper, Christina Nowotny, Silke Michalek, Ulrike Schmidt, and Timo Brockmeyer. 2022. Language style matching and treatment outcome in anorexia nervosa. *European Eating Disorders Review.*

[14] Daniel M Romero, Roderick I Swaab, Brian Uzzi, and Adam D Galinsky. 2015. Mimicry is presidential: linguistic style matching in presidential debates and improved polling numbers. *Personality and Social Psychology Bulletin*, 41, 10, 1311–1319.

[15] A Burcu Bayram and Vivian P Ta. 2019. Diplomatic chameleons: language style matching and agreement in international diplomatic negotiations. *Negotiation and Conflict Management Research*, 12, 1, 23–40.

[16] Lena C Müller-Frommeyer, Niels AM Frommeyer, and Simone Kauffeld. 2019. Introducing rlsm: an integrated metric assessing temporal reciprocity in language style matching. *Behavior Research Methods*, 51, 3, 1343–1359.

[17] Cindy Chung and James W Pennebaker. 2007. The psychological functions of function words. *Social communication*, 1, 343–359.

[18] Debajyoti Pati and Lesa N Lorusso. 2018. How to write a systematic review of the literature. *HERD: Health Environments Research & Design Journal*, 11, 1, 15–30.

[19] Rens Hoegen, Deepali Aneja, Daniel McDuff, and Mary Czerwinski. 2019. An end-to-end conversational style matching agent. In *Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents*, 111–118.

[20] Paul Thomas, Daniel McDuff, Mary Czerwinski, and Nick Craswell. 2020. Expressions of style in information seeking conversation with an agent. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, 1171–1180.

[21] Claes Wohlin. 2014. Guidelines for snowballing in systematic literature studies and a replication in software engineering. In *Proceedings of the 18th international conference on evaluation and assessment in software engineering*, 1–10.

[22] Yla R Tausczik and James W Pennebaker. 2010. The psychological meaning of words: liwc and computerized text analysis methods. *Journal of language and social psychology*, 29, 1, 24–54.

[23] Omer Topaloglu and Mayukh Dass. 2021. The impact of online review content and linguistic style matching on new product sales: the moderating role of review helpfulness. *Decision Sciences*, 52, 3, 749–775.

[24] S Ludwig, KD Ruyter, D Mahr, M Wetzels, E Brueggen, and TD Ruyck. 2014. Take their word for it: the symbolic role of linguistic style matches in user communities. MIS. Q. 38, 1201–1218. doi: 10.25300. Tech. rep. MISQ/2014/38.4.12.

[25] Katharina Heuer, Lena C Müller-Frommeyer, and Simone Kauffeld. 2020. Language matters: the double-edged role of linguistic style matching in work groups. *Small Group Research*, 51, 2, 208–228.

[26] Molly E Ireland and Marlone D Henderson. 2014. Language style matching, engagement, and impasse in negotiations. *Negotiation and conflict management research*, 7, 1, 1–16.

[27] Richard Nordquist. 2020. What are function words in english grammar? (Jan. 2020). https://www.thoughtco.com/function-word-grammar-1690876.

[28] Langtao Chen, Aaron Baird, and Detmar Straub. 2020. A linguistic signaling model of social support exchange in online health communities. *Decision Support Systems*, 130, 113233.

[29] Stephen A Rains. 2016. Language style matching as a predictor of perceived social support in computer-mediated interaction among individuals coping with illness. *Communication Research*, 43, 5, 694–712.

[30] Stephan Ludwig, Ko De Ruyter, Mike Friedman, Elisabeth C Brüggen, Martin Wetzels, and Gerard Pfann. 2013. More than words: the influence of affective content and linguistic style matches in online reviews on conversion rates. *Journal of marketing*, 77, 1, 87–103.

[31] Sarah Peregrine Lord, Elisa Sheng, Zac E Imel, John Baer, and David C Atkins. 2015. More than reflections: empathy in motivational interviewing includes language style synchrony between therapist and client. *Behavior therapy*, 46, 3, 296–303.

[32] Katie Aafjes-van Doorn, John Porcerelli, and Lena Christine Müller-Frommeyer. 2020. Language style matching in psychotherapy: an implicit aspect of alliance. *Journal of Counseling Psychology*, 67, 4, 509.

[33] Jessica L Borelli, Lucas Sohn, BingHuang A Wang, Kajung Hong, Cindy DeCoste, and Nancy E Suchman. 2019. Therapist–client language matching: initial promise as a measure of therapist–client relationship quality. *Psychoanalytic Psychology*, 36, 1, 9.

[34] Amy L Gonzales, Jeffrey T Hancock, and James W Pennebaker. 2010. Language style matching as a predictor of social dynamics in small groups. *Communication Research*, 37, 1, 3–19.

[35] Cathy R Cox and Mike Kersten. 2016. Mortality salience increases language style matching and well-being. *Self and Identity*, 15, 4, 452–467.

[36] Molly E Ireland and James W Pennebaker. 2010. Language style matching in writing: synchrony in essays, correspondence, and poetry. *Journal of personality and social psychology*, 99, 3, 549.

[37] Cristian Danescu-Niculescu-Mizil, Michael Gamon, and Susan Dumais. 2011. Mark my words! linguistic style accommodation in social media. In *Proceedings of the 20th international conference on World wide web*, 745–754.

[38] Katie Aafjes-van Doorn and Lena Müller-Frommeyer. 2020. Reciprocal language style matching in psychotherapy research. *Counselling and Psychotherapy Research*, 20, 3, 449–455.

[39] Elise van der Pol, Sharon Gieske, and Raquel Fernández. 2016. Linguistic style accommodation in disagreements. In *Proceedings of the Fifth Joint Conference on Lexical and Computational Semantics*, 120–124.

[40] Arjun Mukherjee and Bing Liu. 2012. Analysis of linguistic style accommodation in online debates. In *Proceedings of COLING 2012*, 1831–1846.

[41] Simon L Jones, Rachel Cotterill, Nigel Dewdney, Kate Muir, and Adam N Joinson. 2014. Finding zelig in text: a measure for normalising linguistic accommodation. In *Coling*. Vol. 2014, 25th.

[42] Paul Thomas, Mary Czerwinski, Daniel McDuff, Nick Craswell, and Gloria Mark. 2018. Style and alignment in information-seeking conversation. In *Proceedings of the 2018 Conference on Human Information Interaction & Retrieval*, 42–51.

[43] Deborah Tannen et al. 2005. *Conversational style: Analyzing talk among friends.* Oxford University Press.

[44] François Mairesse and Marilyn A Walker. 2010. Towards personality-based user adaptation: psychologically informed stylistic language generation. *User Modeling and User-Adapted Interaction*, 20, 3, 227–278.

[45] Kendra Cherry. 2022. What are the big 5 personality traits? (Oct. 2022). https://www.verywellmind.com/the-big-five-personality-dimensions-2795422.

[46] Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*, 311–318.

A APPENDIX: SEARCH RESULTS

In this section, two tables are given: Table 1 lists the amounts of papers found per category. Table 2 (see next page) lists the search queries that were used and how many relevant papers the queries yielded. In both tables, reference numbers to the papers that are cited in the current article are given.

Category	**Count**	**References**
1	104	[3, 39, 40, 41, 1, 4, 25, 24, 14, 23, 15, 30, 26, 28, 32, 34, 29, 33, 38, 13, 31]
2	18	
3	14	[6]
4	8	[10, 8, 19, 44, 11, 9, 20]
5	6	[12, 16]

Table 1. A table with the count of resulting papers given per each category. The references to the papers that were cited in this article are also given.

Query	**Count**	**References**
linguistic alignment	32	[3, 4, 6, 8, 5],
entrainment	16	[10, 1]
linguistic style matching	20	[16, 12, 28, 26, 30, 24, 15, 23, 14, 25]
linguistic OR language style matching	32	[32, 34, 29, 33, 35, 38, 13, 31, 19, 44, 16, 36]
linguistic style accommodation	9	[39, 40, 41]
linguistic style coordination	4	
linguistic synchrony	1	
adaptive coordination agent	6	
(entrainment OR alignment OR (style matching) OR (style accommodation)) ((human-agent interaction) OR (user adaptation) OR (language generation) OR agent OR chatbot OR (dialogue system))	11	
(dialogue system) OR (conversational (agent OR assistant)) (entrainment OR alignment OR (style matching))	6	[11, 9]
(linguistic OR language) ((style matching) OR (style accommodation)) (human-agent interaction)	5	[20]
((style matching) OR (style accommodation)) ((human-agent interaction) OR (user adaptation) OR (language generation) OR agent OR chatbot OR (dialogue system))	4	

Table 2. A table of the search queries that were used and the amount of resulting papers. The references to the papers that were cited in this article are also given.