Formulaic language of Vietnamese children with autism spectrum disorders: A corpus linguistic analysis

Phạm Hiển
Institute of Linguistics
Vietnam Academy of Social Sciences
phamhieniol@gmail.com

Nguyễn Thị Giang
Institute of Linguistics
Vietnam Academy of Social Sciences
blueriver063@gmail.com

Abstract

This study examines formulaic language (based on multiword expressions) in the interactive speech of eleven children with autism spectrum disorders (ASDs). Play sessions were recorded to collect speech samples. Speech-language pathologists (SLP) acted as informants during the recording sessions. Qualitative and quantitative analyses were carried out: a qualitative analysis of situational factors that potentially impacted the prevalence of formulaic language, a quantitative analysis of the prevalence of formulaic language in speech samples using a classification system developed for the study. Various situational factors increased or decreased formulaic language use, though all eleven participants used formulas. Formulas corresponded to several categories and varied in conventionality, whether in form or function. Nonetheless, the qualitative analysis indicated that formulas had several functional uses in the interactions of participants. These findings have implications for future research and language assessment and intervention in ASD.

1 Introduction

The purpose of this study is to survey the predominance and the essence of the use of formulaic language in the interactions of eleven Vietnamese children with ASD (Nguyễn, 2015). It is situated within the view that language is a complex adaptive system in which language emerges from interactions over time and is formed and entrenched by situational factors encompassing each occurrence of usage. Formulaic sequences are ubiquitous in language use and they make up a large proportion of any discourse. Erman and Warren (2000) calculated that formulaic sequences of various types constituted 58.6% of the spoken English discourse they analyzed and 52.3% of the written discourse. Sinclair's (1991) view that language as a whole is organised according to two main structuring principles: an open choice principle and an idiom principle, with the latter involving the widespread use of formulaic stretches of words. Furthermore, this store of formulaic sequences is dynamic and is constantly changing to meet the needs of the speaker (Wray, 2002: 101).

We have decided to use the term formulaic sequence based on a definition by Wray (2002: 9): "a sequence, continuous or discontinuous, of words or other elements, which is, or appears to be, prefabricated: that is, stored and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar."

This term covers a wide range of formulaic language, and touches on two key criteria of the emphasis in this book: a) we are concerned with sequences of lexis and b) the mind handles, or appears to handle, these sequences at some level of representation as wholes. However, using this definition, Wray argues that even single words and morphemes can be seen as formulaic sequences.

2 Functions of Formulaic Language

As previously described, formulaic language, as opposed to propositional language, is especially well suited to serving certain functions for speakers. Ellis (1996) contended that language acquisition is essentially sequence learning. In fact, children who were late to combine words were more at risk for future problems with language than children who were late with their first words (Rudolph & Leonard, 2016). Wray and Perkins (2000) group the functions of formulaic language into two categories: (a) devices of social interaction and (b) compensatory devices for memory limitations. The first category includes the use of formulas for the purpose of manipulating others, asserting separate identity, and asserting group identity (Wray & Perkins, 2000). Wray (1999: 216) proposes the following scenario: In a crowded and noisy bar, asking a stranger to move so that you can get past requires attracting their attention and interrupting their conversation. A formulaic expression such as *nói chung là* or *thượng lộ bình an* is easily recognized… In contrast, a less formulaic utterance, such as *Tôi đạp xe dắt theo một con chó*, must be heard more accurately because it is unpredictable, and requires more decoding, so it is more intrusive.

3 Functions of formulaic language in ASD.

To date, there appear to be no survey studies of the functions of formulaic language in ASD of Vietnamese children. However, there is reason to believe that formulaic language may serve different functions in persons with ASD as a result of their impairments in social communication (Wray & Perkins, 2000) and difficulties accomplishing full integral perception of all dimensions in communication. Together, these impairments may lead to a situation in which "formulaicity is not socio-interactionally motivated but rather is a 'Hobson's choice' [i.e., having no real alternative] solution to processing constraints" (Wray & Perkins, 2000: 23). In other words, formulaicity in ASD might not socio-interactionally motivated because of the impairment in social communication (Wray & Perkins, 2000) nor may there be a genuine decision between analytic and holistic processing because

difficulties in integrated perception hinder segmentation and thus analytic processing (Noens & Van Berckelaer-Onnes, 2004). With respect to Wray and Perkins' (2000) division in functions, we would thus expect formulaic language to be used for cognitive purposes rather than as social interaction devices. However, impairment in pragmatic abilities in ASD is best described as a *deficiency* than a *complete inability* (Vogindroukas & Zikopoulou, 2011). Indeed, research pertaining to the categories of formulaic language, including immediate and delayed echolalia, politeness sequences (e.g., Volden & Sorenson, 2009), and discourse markers suggest that while the social functions of formulaic language may be impaired or less prevalent, they are not necessarily non-existent. These findings are discussed in greater detail below.

4 Methodology

4.1 Research questions

Q1: Do children on the verbal ASD spectrum with varying language abilities use formulaic language in interactions?
Q2: How are the form and function of formulaic expressions related in the interactions of children on the verbal ASD spectrum with varying language abilities?

4.2 Method

This study approaches the investigation of formulaic language in the conversational speech of children with ASD from a multiple case study design. In a multiple case study, "a number of cases are studied jointly in order to investigate a phenomenon or general condition" (Dörnyei, 2007, p. 152). The cases in this study are children with ASD while the phenomenon of interest is formulaic language use. "Although case studies are typically discussed under the label of qualitative research (because a single case cannot be representative of a population), actual case studies often include quantitative data collection instruments as well" (Dörnyei, 2007, p. 152). Duff (2008) points out that mixed methods data analysis is also appropriate in case studies.

4.3 Participants.

A total of eleven participants (9 males, 2 females) took part in the recording sessions. All participants were ages 36 to 60 months, and had been diagnosed with an autism spectrum disorders by a professional.

Participants were selected using criterion sampling. The key characteristics of interest were age and ASD diagnosis. Additionally, all participants were currently or had previously been clients of our Center for Language Teaching and Rehabilitation, Institute of Linguistics. Not only did this create a point of contact between potential participants and the researcher, but also it ensured that participants would be comfortable interacting with the Center during the play session. No restrictions were made based on sex, but given that ASD affects three to four times more males than females (CDC as cited in Kim & Lord, 2013), it was expected that there would be more male participants.

4.4 Sources of data

For each participant, a one-hour play session with his/her current SLP was audio recorded. The SLP, rather than a parent, was selected as an interlocutor to minimize variations in interpersonal factors across participants. The researchers did not participate as an interlocutor during the play session because previous research has shown that the rate of echolalia tends to increase in interactions with unfamiliar interlocutors. The audio recording device was a Zoom H1 Handy Portable Digital Recorder. The audio recording settings were: WAV at 96Hz at 16Bit.

4.5 Participant Profiles

This section provides background information for each of the participants as well as a brief description of their recording session.

Table 1: *Summary of Demographic Information for Participants*

Pseudonym	Sex	Age at time of recording (months)	Age at ASD diagnosis (months)
T01	Male	37	36
T02	Female	39	30
T03	Male	41	38
T04	Male	44	42
T06	Male	47	40
T07	Male	50	48
T08	Male	53	36
T09	Male	54	48
T10	Male	55	54
T11	Female	59	54

4.6 Quantitative Analysis of Formulaic Language

This section provides the results of a quantitative analysis of formulas from each participant's transcript. This section presents and discusses: (a) the distribution of formulaic and non-formulaic speech by word count, (b) the distribution of formulaic expressions across categories, (c) the distribution of formulaic expressions by function, and (d) the variability of formulaic expressions. Excerpts selected for quantitative analysis were chosen randomly to prevent biases that would confirm or disconfirm the researcher's expectations. They ranged from 4000 to 11000 words each. Word counts were based exclusively on participants' utterances. Coded transcripts for the selected segments are as following Figure 1, along with a description of the ongoing activity. Each coded transcript is followed by a list of formulaic expressions organized by category.

Figure 1: Excerpt concordance of the transcripts

4.7 Distribution of formulaic and novel language

As the actual length of excerpts varied at length, the distribution by word count of formulaic and novel language between participants was compared using percentage scores (WC%) as opposed to raw scores. Figure 2 illustrates the overall distribution

of novel and formulaic language, which has been subdivided into unconventional verbal behavior (UVB) formulas, and conventional formulas. Unconventional verbal behavior formulas include immediate and delayed echolalia, as well as perseveration. Conventional formulas include all other types of formulaic expressions.

Based on the data in Figure 2, the total WC% of formulaic language, including both UVB and conventional formulas, varied between 30% and 80% in the excerpts selected. Thus, while all speakers used formulaic language, they did not use it to the same extent. Furthermore, the figure above indicates that T02 did not use any formulas that were classified as UVB in the excerpt selected for analysis. According to Wray's (2002) model of the balance of holistic and analytic processing, it would be expected that the older participants, T02 and T04, would use more formulaic language than the younger participants, T01 and T04. However, Figure 2 illustrates that this was not the case in the excerpts selected for analysis. T04 used the most formulaic language as measured by WC%, but T02 used the least of all eleven participants. This variability across speakers is not surprising as the excerpts for analysis were randomly selected. Thus, the contextual factors for each segment varied considerably.
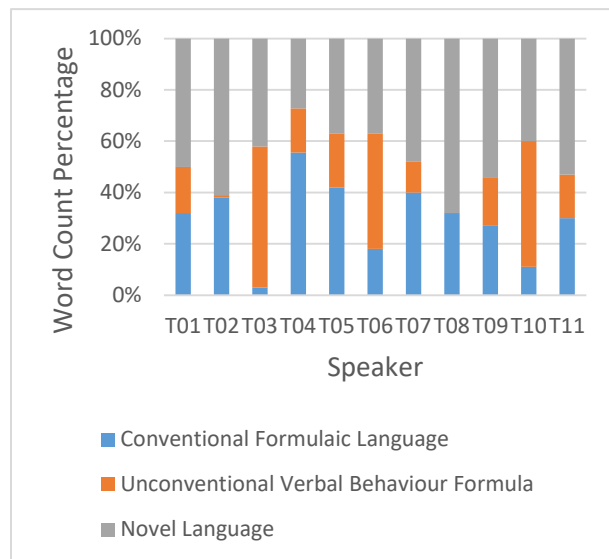

Figure 2. Distribution by word count percentage of novel and formulaic language, by speaker.

As discussed in the previous section, contextual factors influence the prevalence of formulaic language in conversational speech. In terms of ongoing activity, the participants engaged in a number of activities in the segments, including playing with trains, animal figures, and puppets, and participating in conversation without any play. The cognitive and conversational demands placed on the participants in each segment varied according to the type of play, and so the distribution WC% of formulaic language likely reflects these differences in situational factors. If more segments had been analyzed for each participant, age trends may have been clearer. Although there is a relationship between overall number of formulaic expressions and WC% of formulaic and novel language, this measure of formulaicity does not distinguish between the length of sequences and their frequency. By measuring formulaic language use according to WC%, the length of formulas may be a confounding factor in the relationship between formulaic language use and age. Discrepancies such as the difference between T03 and T04 in their respective WC% of conventional formulas are highly related to the length of sequences. While T04's conventional sequences included a number of phrase-long exemplars, T03's longest conventional sequence in the excerpt was 5 words long, while the rest were 3 words or less. The frequency of sequences and its relationship to WC% of formulaic and novel language is addressed later. Another aspect of the distribution of WC% of formulaic and novel language that was of particular interest was the WC% of unconventional verbal behavior compared with all other types of speech. In the Introduction section, an alternate explanation was proposed for Van Lancker Sidtis's (2012) observation-based account of formulaic language in ASD. Van Lancker (2012) put forward the notion that the speech of high-functioning persons with ASD is lacking in formulaic language while the speech of low-functioning persons with ASD is high in formulaic language. However, in school-aged children, the distinction between high-functioning and low functioning autism in school-aged children is based on expressive language abilities, such that strong expressive language is associated with high-functioning autism. Therefore, expressive language abilities can be used to approximate the distinction between high- and low-functioning autism made by Van Lancker Sidtis (2012). Van Lancker

Sidtis's (2012) observations regarding the prevalence of formulaic language do not align with previous research findings, which indicate that formulaic language appears to be characteristic of the entire verbal ASD spectrum. Therefore, it was suggested that Van Lancker Sidtis's (2012) impressions were potentially the result of equating formulaic language with UVB and novel language with all other speech, including conventional formulas. These operational definitions have been observed elsewhere in ASD research. Thus, according to a revised set of definitions, it was proposed that Van Lancker Sidtis (2012) in reality was observing that UVB, and not formulaic language as a whole, decreases with an increase in expressive language abilities. As shown in Figure 2, all participants in this study, regardless of language abilities, used formulaic language. Thus, it is worth investigating the hypothesis discussed above and in the Introduction to attempt to reconcile conflicting accounts of formulaicity in ASD. As the distinction between high-functioning and low-functioning autism in school-aged children is based on expressive language (Tager-Flusberg et al., 2005), participants were ranked in terms of their expressive language abilities to approximate the distinction made between persons with ASD using the previously-mentioned labels. Using this ranking, it was possible to determine whether the participants' WC% of novel and formulaic language as defined in formulaic language research coincided with Van Lancker Sidtis's (2012) observations. If they did not, then the proposal that the observations were based on different operational definitions of formulaic and novel language could be tested. Participants were initially ranked according to their expressive language abilities based on the Child participant profile questionnaire. Based on these measures, T06 had the strongest expressive language skills, as his mother indicated he never has poor expressive language and uses complex sentences. Conversely, T07 had the weakest expressive language skills based on these measures, as his mother indicated that he uses one-word utterances and frequently has poor expressive language. T09 and T10 were more closely matched; while T09 uses compound sentences and T04 uses complex sentences, T10 always has poor expressive language while T09 occasionally does. Thus, based on the questionnaire items, the participants were ranked from strongest to least strong expressive language skills as follows: T03, T06, T09, T10, T11, T01, T02, T04, T05, T08, and T7. The SLP was also asked to rank the participants according to their expressive language abilities based on her professional experience working with the participants and her observations during the recording sessions. Her assessment agreed with the ranking established using the Child participant profile questionnaire. She also indicated that T01 and T04 were difficult to rank as they have different strengths. Therefore, it is with a certain degree of confidence that we can rank the participants based on their expressive language abilities as follows: the strongest is T03, followed by, T01, T04, and T03. With this ranking established, the first step in testing the application of Van Lancker Sidtis's (2012) observations in this study was to compare the ranking by expressive language abilities with the ranking by WC% of formulaic language presented in Figure 2. Based on WC% of formulaic language, T03 used the highest proportion of words in formulaic expressions, followed by. T06, T09, T10, T11. Conversely, according to Van Lancker Sidtis's (2012) observations and the participants' expressive language skills, they should have been ranked according to their WC% of formulaic language as follows: T03, T06, T09, T10, T11, T01, T02, T04, T05, T08, and T7. As the expected ranking and the actual ranking of participants based on WC% of formulaic language did not coincide, it was concluded that the use of formulaic language was moderated by factors other than or in addition to expressive language abilities. As participants' use of formulaic language did not coincide with Van Lancker Idtis's (2012) account of formulaic language in ASD, the proposal that the observations were based on different operational definitions of formulaic and novel language was tested.

4.8 Summary of Findings

This study used several approaches to examine the prevalence and nature of formulaic language use in the interactions of eleven Vietnamese children with ASD. This section summarizes the findings and discussions of the context factors analysis, the quantitative analysis of formulaic language, and the qualitative analysis of formulaic sequences.

The summary is organized in relation to the research questions that guided this study.

The research questions presented in the previous section are repeated below for the purpose of reviewing the issues that were of interest in this study.

Question 1. Do children on the verbal ASD spectrum with varying language abilities use formulaic language in interactions?

This study found that children on the verbal ASD spectrum with varying language abilities did use formulaic language in conversation. The overall balance of formulaic to novel language varied from participant to participant. The percentage of words in formulaic expressions ranged from 30% to 80% of the total word count of each participant's excerpt. This variation is not surprising given the various situational factors at play. While some factors impacted the language system more globally, others had a more direct impact on language use at the moment of speech.

Question 2. How are the form and function of formulaic expressions related in the interactions of children on the verbal ASD spectrum with varying language abilities?

Various form-function combinations of formulas were observed in the speech of the participants. Functional uses of formulas were expected based on results of the quantitative analysis that indicated that all eleven participants used more formulas associated with pragmatic functions than formulas without a pragmatic function or with no function whatsoever. Nonetheless, a qualitative analysis was required to confirm the relationship between form and functional uses of formulaic sequences because of the quantitative analysis's focus on identification and classification by form. In the quantitative analysis only immediate and delayed echolalia were analyzed in terms of function. The functions of all other formulas were implicit based on their respective categories. For example, all collocations belonged to the group of formulas that had no pragmatic functions while all conventional expressions belonged to a subset of formulas that did have a pragmatic function.

5 Conclusions

The purpose of this multiple case study was to examine both the prevalence and the nature of formulaic sequence use in the interactions of eleven children with ASD within the framework of language as a complex adaptive system. To this effect, the participants took part in a one-hour audio-recorded play session with their speech-language pathologist to gather language samples. The participants' parents acted as informants while observing the play session. The audio recordings and the information provided by the parents were used to create participant profiles and analyze the situational factors surrounding the participants' use of language. The researchers transcribed the recordings in order to carry out quantitative analyses of the prevalence and qualitative analyses of the natures of use of formulaic language. The main findings of the study pertain to three domains: situational factors, the prevalence of formulaic expressions, and the nature of use of formulaic expressions. This study found evidence that several situational factors impacted the participants' language use and that these factors did not act in isolation but rather interacted together. The results of the quantitative analyses indicated that all participants, regardless of expressive ability, used formulaic language. All eleven participants used conventional sequences. Conversely only nine of the eleven participants used unconventional sequences characteristic of disordered language, including immediate and delayed echolalia, and perseveration. All eleven participants used sequences from a range of categories of formulaic language and additionally used varied sequences, such that they were more likely to use a new formula than to repeat an old one. Finally, a qualitative analysis of the nature of use of formulaic language was carried out on 36 exemplars selected throughout the transcripts. The sequences were analyzed in relation to the surrounding speech and ongoing activity to determine the markers of formulaicity associated with each and the sequences' functions in context. The sequences were categorized as belonging to one of three form-function pairings: idiosyncratic formulas with functions, conventional formulas with idiosyncratic functions, and conventional formulas with conventional functions. Participants

used the formulas for a range of functional purposes, both interactive and non-interactive.

References

Dörnyei, Z. (2007). *Research Methods in Applied Linguistics*. Oxford, UK: Oxford University Press.

Duff, P. A. (2008). *Case study research in applied linguistics*. New York, NY: Lawrence Erlbaum Associates.

Ellis, N. C. (1996). Sequencing in SLA: Phonological memory, chunking, and points of order. *Studies in Second Language Acquisition*, 18, 91–126.

Erman, B. and Warren, B. 2000. The idiom principle and the open-choice principle. *Text* 20: 29–62.

Kim, S. H., & Lord, C. (2013). The behavioral manifestations of autism spectrum disorders. In J. D. Buxbaum & P. R. Hof (Eds.). *Neuroscience of autism spectrum disorders* (pp. 25-37). Oxford, UK: Academic Press.

Nattinger, J. R. and DeCarrico, J. S. 1992. *Lexical Phrases and Language Teaching*. Oxford: OUP.

Nguyễn Thị Hoàng Yến, (2015). *Tự kỉ: Những vấn đề lí luận và thực tiễn*. Education Publishing House.

Noens, I., & Van Bercklaer-Onnes, I. (2004). Making sense in a fragmentary world: Communication in people with autism and learning disability. *Autism*, *8*(2), 197- 218.

Rudolph, J. M. & Leonard, L. B. (2016). Early language milestones and specific language impairment. *Journal of Early Intervention*, 38(1) 41 –58.

Sinclair, J. M. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.

Tager-Flusberg, H., Paul, R., Lord, C. (2005). Language and communication in autism. In F. R. Volkmar, R. Paul, A. Klin, & D. Cohen (Eds.), *Handbook of autism and pervasive developmental disorders* (3rd Ed., Vol. 1, pp. 335-364). Hoboken, NJ: John Wiley & Sons.

Van Lancker Sidtis, D. (2012). Formulaic language and language disorders. *Annual Review of Applied Linguistics*, *32*, 62-80.

Vogindroukas, I., & Zikopoulou, O. (2011). Idiom understanding in people with Asperger syndrome/high functioning autism. *Revista da Sociedade Brasileira de Fonoaudiologia*, 16*(4), 390-395.*

Volden, J., & Sorenson, A. (2009). Bossy and nice requests: Varying language register in speakers with autism spectrum disorder (ASD). *Journal of Communication Disorders*, *42*(1), 58-73.

Vũ Thị Bích Hạnh, (2019) *Đối mặt với tự kỉ: Cùng nhau vượt qua.* Tri Thuc Tre books, Women Publishing House.

Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge, UK: Cambridge University Press.

Wray, A., & Perkins, M. R. (2000). The functions of formulaic language: An integrated model. *Language & Communication*, *20*(1), 1-28.