



Preferred Argument Structure: A Comparative Analysis of Arabic Discourse

Essa Alfaihi*

Department of Arabic language and literature, Jazan University, Jazan, Saudi Arabia

Received: 9/5/2023

Revised: 20/7/2023

Accepted: 27/9/2023

Published online: 27/8/2024

* Corresponding author:

ealfaihi@jazanu.edu.sa

Citation: Alfaihi, E. . (2024).

Preferred Argument Structure: A

Comparative Analysis of Arabic

Discourse. *Dirasat: Human and*

Social Sciences, 51(5), 464–478.

<https://doi.org/10.35516/hum.v51i5.4>

627

Abstract

Objectives: This paper offers a new analysis procedure of Arabic discourse by examining the correlation between information flow and grammatical forms, specifically the core arguments of the verb: the subject and the direct object.

Methods: The discourse analysis in this paper compares three types of spoken texts which differ across time and genre: Spontaneous spoken Arabic, spontaneous spoken Standard Arabic, and spoken Classical Arabic. The analysis is within the framework of Preferred Argument Structure (PAS) which refers to the observed tendency for speakers to avoid expressing more than one lexical argument or more than one piece of new information in a clause.

Results: The results show that information flow in Arabic discourse is generally consistent with the cognitive constraints of PAS, although the results varied according to the type of discourse. Contemporary spoken Arabic exhibited the most consistency with PAS whereas Classical Arabic exhibited the least consistency.

Conclusions: The distribution of information across the core arguments of the verb forms patterns that can be statistically tracked and predicted. The subject of the transitive verb is for light information that does not require great cognitive effort among speakers, such as pronouns or familiar information. In comparison, the object and subject of the intransitive verb are for information that require greater cognitive effort like full nouns or new information. The study recommends applying PAS in future studies since it is applicable to multiple aspects of linguistic studies such as language acquisition, discourse analysis, and comparison between spoken and written languages.

Keywords: Information flow, discourse analysis, activation cost.

البنية المفضلة في الجملة: تحليل مقارن للخطاب العربي

عيسى الفيضي*

قسم اللغة العربية وآدابها، جامعة جازان، جازان، المملكة العربية السعودية

ملخص

الأهداف: يعرض هذا البحث آلية جديدة لتحليل الخطاب العربي عن طريق دراسة المعلومات في الخطاب وتوزيعها على الأركان الرئيسية في الجملة كالفاعل والمفعول به.

المنهجية: يعقد تحليل الخطاب في هذه الورقة مقارنة بين ثلاثة أنواع من النصوص المنطوقة التي تختلف في نوع الخطاب وزمنه، وهي: خطبة منتقاة من عربية التراث، وحوار في العربية الفصحى المعاصرة، ومحادثة بإحدى اللهجات العربية. ويدور التحليل في إطار منهج "البنية المفضلة في الجملة" (PAS by Du Bois, 2003) وهو المنهج الذي يفسر تفاد المتحدثين (بأي لغة كانت) استخدام أكثر من حجة لفظية جديدة في الجملة الواحدة.

النتائج: أظهرت نتائج البحث أن توزيع المعلومات في الخطاب العربي كانت متسقة على نحو عام مع "القيود الذهنية" تحت منهج "بنية الحجج المفضلة في الخطاب"، مع اختلاف طفيف وفقاً لنوع الخطاب. فقد أظهر التحليل أن اللهجة العربية كانت الأكثر اتساقاً مع "بنية الحجج المفضلة" أما عربية التراث فكانت الأقل اتساقاً.

خلاصة الدراسة: خلصت الدراسة إلى أن توزيع المعلومات على أركان الجملة يشكل أنماطاً يمكن تتبعها إحصائياً والتنبؤ بها. فالفاعل في جملة الفعل المتعدي هو محل للمعلومات التي لا تتطلب جهداً ذهنياً كبيراً بين المتكلم والمخاطب كالضمائر أو المعلومات المألوفة للمتخاطبين. أما المفعول به وفاعل الفعل اللازم فهما محل للمعلومات التي تتطلب جهداً ذهنياً أكبر كالأسماء أو المعلومات غير المألوفة للمتخاطبين. وتوصي الدراسة بأهمية تطبيق المنهج التحليلي لـ "بنية الحجج المفضلة" في الدراسات المستقبلية كونه منهجاً قابلاً للتطبيق على جوانب متعددة من الدراسات اللغوية كالتساب اللغوية، وتحليل الخطاب، والمقارنة بين لغتي النطق والكتابة.

الكلمات الدالة: توزيع المعلومات، تحليل الخطاب، الجهد الذهني لتنشيط المعلومات.



© 2024 DSR Publishers/ The University of Jordan.

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY-NC) license <https://creativecommons.org/licenses/by-nc/4.0/>

1. Introduction

If we think of the purpose of communication (any communication), we will find that the main purpose of communication is to exchange information, whether in casual conversations or formal speech events. Spoken language does not always constitute a single form of discourse. Some forms of spoken language are formal, highly technical, and involve complex grammatical structures and advanced vocabulary (e.g., political speeches, and scientific lectures), whereas other types of spoken language have relatively simple grammatical structures (e.g., spontaneous spoken language at home or between friends). However, if the purpose of all communications is to exchange information, then the complexity vs. simplicity of different registers of spoken interaction has to do with how information is handled in a particular interaction.

There are different approaches that examine information management in discourse, but the one that this study is concerned with is Du Bois' approach (1987, 2003). Through his intensive study in 1987 of Sacapultec narratives (a Mayan language spoken by about 10,000 people in the highlands of Guatemala), Du Bois had a fascinating observation of how language users go about conveying information in their discourse. One of the most compelling conclusions of his study was the *Preferred Argument Structure* (hereafter referred to as PAS), which can be defined as a set of empirical constraints on the information status and coding of direct arguments in a clause. The main claim of Du Bois's study is that there are measurable statistical tendencies regarding how the argument structure of a given verb is realized in a particular discourse (Mittelberg, 2007).

1.1 Significance of PAS

PAS offers a unique method to analyze and understand how noun phrases connect with their verbs in a discourse since it combines different aspects of language, such as semantics, syntax, morphology, typology, diachrony, acquisition, cognition, and pragmatics (Du Bois, 2003). As the analysis in the present paper will demonstrate, PAS track and code every noun phrase in a discourse for its grammatical properties (i.e., whether the clause is transitive, intransitive or verbless), morphological properties (i.e., whether the noun phrase is pronouns, lexical, or null), semantic, cognitive and pragmatic properties by tracking the activation process of the information (i.e., whether the information is old, new or somewhat in between old and new).

This coding procedure of every clause in a discourse gives a detailed statistical description about the discourse that can be applied to various discourse investigations whether it is typological, diachronic, acquisition, etc. PAS approach looks inside and outside the clause by investigating the grammatical forms and patterns of a discourse and bringing it together with the large pragmatic context. Before going into the details regarding PAS, I shall briefly introduce the *activation-cost* approach by Chafe (1994) since it contains a few important terminological preliminaries.

The remaining parts of this paper are organized as follows: Section 2 introduces the theoretical framework including the Activation Cost Hypothesis and PAS. Section 3 contains the objectives of the present study and research questions. Section 4 covers previous literature on the topic. Section 5 includes the methodology and coding criteria. Section 6 presents the results. Section 7 includes the discussion, and Section 8 contains the concluding remarks.

2. Theoretical Framework

2.1 Activation Cost Hypothesis

Chafe (1994) examined the nature of information in discourse. He hypothesized that when a speaker expresses a particular concept in discourse, the concept may be in any one of three different states of activation. He uses the terms active, semi-active, and inactive and defined these terms as follows:

“An **active** concept is one that is currently lit up, a concept in a person's focus of consciousness. A **semi-active** concept is one that is in a person's peripheral consciousness, a concept of which a person has background awareness, but which is not being directly focused on. An **inactive** concept is one which is currently in a person's long-term memory neither focally nor peripherally active” (1994, p.25).

To illustrate these terms, suppose we have two college students having a normal conversation about a course they are

taking together:

- Example (1) Student A: I saw the professor and she answered my question.
Student B: Really! That is great.
- Example (2) (Students talking about a textbook)
Student A: so did you buy it?
Student B: I haven't, but I will soon.
- Example (3): Student A: I was searching for references and a nice librarian offered to help me.
Student B: That was nice of her.

Now, consider the way speaker A introduced the following referents in the previous three examples: *the professor* in example (1), the singular pronoun *it* in example (2), and *a nice librarian* in example (3). Obviously, speaker A had three different assumptions about the status of those referents in the mind of speaker B when he uttered those concepts. The speaker-hearer schema makes the difference here. In example (3) (*a nice librarian offered to help me*), the entire event of offering help was in some sense new within this particular discourse; hence the speaker anticipated that this event was totally *inactive* in the hearer's mind so that he chose to express this *inactive* information in a full noun phrase (a nice librarian). The opposite happened in example (2) (Did you buy it) in which the referent of the pronoun *it* (i.e., the textbook) was already as well known to the listener as to the speaker; hence, the speaker easily anticipated that this piece information was totally *active* in the hearer's mind. The case is different in example (1). The conversation was about a particular course that the two students were taking together, and there was only one professor teaching that course; hence, this particular referent (the professor) cannot be said to be totally active or totally inactive in the listener's mind, but rather somehow *semi-active*. It depends on the recoverability of the hearer and how the listener judges the hearer's consciousness. Hence, the speaker in this hypothetical example anticipated that the hearer might not recover the referent if he said, for example (I saw her, and she answered) instead of (I saw the professor, and she answered).

According to Chafe (1994), it is helpful to think of these three processes in terms of cognitive cost: active information is the least costly, semi-active information is somewhat more costly, and inactive information is the most costly.

Chafe then associated the terms *given*, *accessible*, and *new* to the concepts that were mentioned above; active, semi-active, and inactive, respectively. He defined a **given** information in discourse as one "that the speaker thought was already known to the listener", and **new** information as "an idea that the speaker thought was previously unknown to the listener" (1994, p.71). **Accessible** information in discourse remains in status between newness and givenness and tends to be expressed in more or less the same way as new.

While the concepts of *given* and *new* are clear in Chafe's definition, *accessible* is somewhat vague. Therefore, Prince's (1981) definition of what he called "inferable information" better explains Chafe's accessible information. *Inferable* information is when the speaker assumes the hearer can infer the referent, via logical or plausible reasoning, from discourse entities already evoked. For example, if a speaker said 'John's brother worked with me', then the referent (John's brother) would be considered as *given* (if John has only one brother) or *accessible* (if John has more than one brother). Therefore, if the hearer responded immediately with a question like (which brother?) Or with a statement like (oh, I like working with that guy), then the discourse analysts determine that the referent of (John's brother) was accessible in the former case, and given in the latter.

In light of Chafe's taxonomy of activation costs (given, new, accessible), Du Bois (1987) has proposed the PAS framework that has both a grammatical dimension and a pragmatic dimension. The following section discusses it in detail.

2.2 Preferred argument structure (PAS)

In normal communication, speakers produce several possible forms of arguments. The arguments that this paper is concerned with are the core arguments of the verb in an independent clause: the subject of an intransitive verb (conventionally symbolized as S), the subject of a transitive verb (conventionally symbolized as A), and the direct object

of a transitive verb (conventionally symbolized as O). Arabic indeed has an additional core argument that occurs in verbless clauses, but I will deal with it later in the methodology and coding section (Section 5).

Regarding the form of an argument, speakers may select one of three possible forms (lexical, pronominal, or null) to express the information and fill in the core arguments of the verb. However, Du Bois observed that the speaker's selection of verb arguments is governed and motivated by the PAS. PAS is a set of empirical constraints on the information status and coding of direct arguments in the clause. These constraints may be stated as follows:

1. **Avoid lexical A** (so-called non-lexical A constraint).
2. **Avoid new A** (so-called given A Constraint).
3. **Avoid two or more lexical arguments per clause** (so-called one lexical argument constraint).
4. **Avoid two or more new arguments per clause** (so-called one new argument constraint).

Although many scholars call them constraints, they, in fact, are rather like preferences for a particular discourse pattern, which speakers are free to violate, but evidently, speakers rarely do (Genetii & Crain, 2003). To illustrate, consider the following examples:

Example 4. *He opened it.*

Example 5. *He opened the door.*

Example 6. *John opened the door.*

Example 7. *A guy opened the door.*

Example 8. *A guy opened a door.*

Although all these sentences are grammatically correct and any of them can be used at any time by the speaker, PAS posits that sentences (4) and (5) should occur naturally in any spontaneous language more frequently than sentence (6); and sentence (6) occurs more than sentence (7). However, sentence (8) occurs rarely if ever. Let us consider each sentence according to the PAS constraints.

In sentence 4, (he opened it), the A argument is evoked by the pronoun *he* (hence A is *pronoun* and *given*), and O argument is evoked by the pronoun *it* (hence O is *pronoun* and *given*). Therefore, all four constraints are met (i.e., constraint 1- Avoid lexical A, 2- Avoid new A, 3- Avoid more than one lexical argument per clause, and 4- Avoid more than one new argument per clause).

In sentence 5, (He opened the door), the A argument is *pronoun* and *given*, but the O argument is invoked by a full noun phrase that is still given; hence O is *lexical* and *given*. Therefore, all four constraints are met in sentence (5).

In sentence 6, (John opened the door), the A argument is *lexical* and *given*, and the O argument is *lexical* and *given* as well. This sentence violates constraint (1) because it has a lexical A and violates constraint (3) because both argument positions are filled with full noun phrases.

In sentence 7, (A guy opened the door), the A argument is *lexical* and *new* whereas O is *lexical* and *given*. Therefore, this sentence violates three constraints: (1) and (2) because the A role is lexical and new, and it also violates constraint (3) because both A and O roles are lexical.

In sentence 8, (A guy opened a door), the A argument is *lexical* and *new*, and the O argument is *lexical* and *new*. Therefore, this sentence violates all four constraints because both arguments are lexical and new. According to the PAS framework, such a sentence occurs rarely if ever in spontaneous speech.

Regarding universality, PAS posits that speakers of any language produce several possible forms of arguments. However, their preference when they code referential forms into the verb's core arguments is believed to be governed by their awareness of the interlocutor's cognitive status. In this sense, PAS describes universally similar cognitive costs of activating and tracking referents in discourse. Du Bois puts it as "When attested spontaneous language use is systematically investigated, there will be consistent statistical tendencies marked by the predominance of certain grammatical configuration of argument realization, and the relative scarcity of others" (Du Bois, 2003, p.34).

3. The present study

The study of language and consciousness together is an opportunity to gain insight into how the mind processes

language. PAS provides a window into how language users manage the information flows in spontaneous language. The Arabic diglossic situation offers a unique perspective on examining PAS constraints among different varieties of Arabic. The present study aims at analyzing three different Arabic discourse genres: A spontaneous spoken dialect, a spontaneous spoken Modern Standard Arabic (hereafter referred to as MSA), and what is supposed to be a spoken text (speech or khutbah) of Classical Arabic.

The examination of the three varieties of Arabic provides insight into the universality of PAS constraints across language genres (i.e., formal vs. informal Arabic) and across time (i.e., Classical Arabic vs. contemporary Arabic data). This study attempts to use the PAS discourse analysis toolbox to look for consistent patterns and tendencies of argument distribution in Arabic discourse cross-language genres as well as cross-time. The study attempts to answer the following two research questions:

Q.1 Does Arabic discourse (old Classical Arabic, MSA, and the colloquial form) display a preferred argument structure?

Q.2 Based on discourse analysis of the three varieties of Arabic, what are some similarities and differences (if any) with respect to information management and PAS constraints?

The following section outlines some of the previous work on the topic with an emphasis on a study by (Ashby & Bentivoglio, 2003) whose research question about Spanish and French discourse motivated the question regarding the Classical Arabic discourse portion of this research.

4. Literature Review

Although PAS principles were first derived and analyzed using narrative data from Sakapultek Maya (Du Bois, 1987), subsequent research has made a great contribution to the universality of PAS by analyzing different genres in a variety of languages, such as English (Bloom, 1990; Kohn & Cragnolino, 2003; Kumpf, 2003), Chinese (Chui, 1992; Huang, 2012; Lin 2009), Inuktitut (Allen, 2000; Allen, & Schröder, 2003), Portuguese (Everett, 2009), Italian (Serratrice, 2002 as cited in Allen, 2007), Japanese (Matsumoto, 2000), Tsou (Huang & Huang, 2009), French (Ashby & Bentivoglio, 1993), Spanish (Ashby & Bentivoglio, 1993), Korean (Clancy, 2003), Nepali (Genetii & Crain, 2003), Tondano (Brickell & Schnell, 2017), among others. However, while those scholars used the principles of PAS as their theoretical framework, most of them moved beyond it to explore some new directions. Some investigated the developmental origins of PAS in children's discourse (Bloom, 1990; Clancy, 2003; Allen, 2000; 2007; Huang, 2012; Serratrice, 2002). Others examined different topics such as classroom interaction (Kumpf, 2003), Autism (Weber, 2003), and children with specific language impairment (Jiang et al, 2022). The overall findings of previous studies were quite consistent in support of the universality of PAS.

However, although much research has reported that PAS seemed to be a powerful phenomenon across languages and genres, it is rarely tested across time, except in one study by Ashby & Bentivoglio, (2003) in which they tested the PAS hypothesis on Old French and Old Spanish texts. They compared the results of those old texts to modern languages (particularly modern Spanish and French) to determine whether changes have occurred across time (i.e., old and modern texts) and space (i.e., France and Latin America). The medieval epics they analyzed exhibited the characteristics of PAS such as the A role was disfavored for encoding new referents and a low incidence of lexical NP overall the data. In addition, the data strongly supported both *one lexical argument* and *one new argument* constraints.

The present study attempts to use a similar approach with Arabic discourse including three varieties of Classical Arabic, MSA, and colloquial Arabic to see how information is encoded among the three different varieties of Arabic discourse.

To the best of my knowledge, the four constraints of PAS have not been examined in Arabic discourse to see whether Arabic varieties (Classical Arabic, MSA, and colloquial) would exhibit consistency with PAS constraints. It is important to note that some previous work has addressed the topic of information structure in Arabic such as the work by Owens et al. (2010) and Holes (2004). However, while their work revolves around information management in Arabic clauses, it did not address PAS constraints. Owens et al. (2010) examined in detail the overt vs. covert subjects in Arabic clauses in spoken Emirati, Kuwaiti, and Hijazi Arabic whereas Holes (2004) discussed principles of weight and rhythm in Standard Arabic clauses.

5. Methodology

5.1 Data

There were three different Arabic oral data selected for the analysis:

5.1.1 Classical Arabic data (hereafter CA)

Since the oral data from Classical Arabic are obviously unavailable, CA data were instead obtained from a written version of spoken CA at that time. The data were obtained from the collection of Al-Jahiz (776- 869AD; see Haruin, 1968). These spoken texts are supposed to be performed spontaneously at that time, but this claim must be taken with caution. The CA data came to us in a written form, and we cannot be sure whether the author edited those texts and, if so, to what extent they were edited. In fact, PAS makes a perfect instrument in analyzing such texts from old spoken language because it will help the analyst to navigate through the constructions and information management in medieval texts and compare them with the contemporary spoken language. Such an approach is believed to help determine whether the clause structure in medieval texts would show more characteristics of spoken-like or written-like language.

5.1.2 Modern Standard Arabic (hereafter MSA).

The data of MSA were collected from a spontaneous conversation (live TV interview) with Mahmoud Darwish (1941-2008), a famous poet and author in the Arabic world. The interviewer was Nashwa Al-Ruwaini, an Arabic news anchor and TV figure. The interview was live in a studio in front of around 15 people as the audience who occasionally asked questions. The video was obtained from YouTube (Startpag, 2012).

5.1.3 Faiḍi Dialect (hereafter FD).

FD is a dialect of Arabic spoken in the southwest region of Saudi Arabia. The data of FD were obtained from a spontaneous conversation recorded between two native speakers of FD for the purpose of this study. The conversation topic was the experience of traveling abroad.

In both FD and MSA, one speaker performed most of the talking with little interruption/overlaps except for a few instances in which the addressee asked a short question or offered backchannels in the conversation. As for CA data, one cannot be sure how it occurred, but as it appears in Al-Jahiz's collections, they were delivered as a speech in front of a public audience (Khutbah) so the assumption is that the interruptions/overlaps were limited, if any.

5.2 Coding

The coding followed the coding procedure in most of the earlier studies on PAS where core arguments in each clause were defined and coded for three types of arguments and functions (see Table 1):

- 1) Coding for syntactic roles whether A, S, Sn, or O.
- 2) Coding for form whether L, P, or Zero.
- 3) Coding for activation state whether N, G, or ACC,

Table 1 below shows the coding key points and definitions. This coding procedure of clause arguments follows the taxonomy by Chafe (1994) and Du Bois (2003).

Table 1. Factors coded for each clause core and definitions.

Factor		Definition
Syntactic role	A	Subject of a transitive verb
	S	Subject of an intransitive verb
	Sn	Subject in verbless sentences
	O	Object of a transitive verb
Form	L	Lexical (one word or juxtaposition)
	P	Pronominal
	Zero	Null argument
Activation	N	New
	G	Given

Factor		Definition
	Acc	Accessible

Moreover, to have an equivalent amount of data, one hundred main clauses from each type of data were selected for the analysis. These are the first one hundred main clauses from the beginning of each text. Following the tradition of previous studies on the topic, the subordinate clauses were included in the analysis too, resulting in a total of 398 clauses which are included in the analysis as well. The distribution of clauses among the three varieties is shown in Table 2 below.

Table 2. Distribution of clause types among the three data

	CA	MSA	FD
Main	100	100	100
Subordinate	39	46	13
Total	139	146	113

The following are three examples that were selected from each speech data to illustrate the coding procedure:

Example from FD:

Clause: [sakan-t fi ʕarʕa]

Gloss: lived- 1sg in Arar

English: I lived in Arar (city).

This FD clause is coded as follows: role = S, form = P, activation = G, which translates to the subject of the intransitive verb is pronominal and given.

Example from CA:

Clause: [guluub-u-hum qariha]

Gloss: heart-NOM-3mpl sore

English: Their hearts are sore.

This CA clause is coded as follows: role = Sn, form = L, activation = G, which translates to the subject of the verbless sentence is lexical and given.

Example from MSA:

Clause: [taxlug xuʕuuba fii al-kitaaba al-ʕifriya]

Gloss: creates.3fsg fertility in the-written the-poetry

English: It creates fertility in the poetry.

This MSA clause is coded as follows: role = A, form = zero, activation = G, role = O, form = L, activation = N, which translates to the subject of the transitive verb is zero (null) and given; and the object is lexical and new.

6. Quantitative Results

PAS posits that spontaneous language production should show consistent statistical tendencies towards certain grammatical and pragmatic roles and quantities, particularly A and O roles. The results of the present analysis of the three texts generally reveal a distribution of forms that are consistent with PAS constraints. However, there were differences among the three varieties of Arabic in their conformation to the four constraints of PAS. FD confirms the most, followed by MSA, followed by CA. The remainder of the section is devoted to presenting the quantitative results on each constraint individually considering the three varieties together. Then, a discussion section is presented including a qualitative interpretation along with illustrative examples from the data.

Constraint 1. Avoid more than one lexical argument per clause.

The focus here is on the proportion of lexical noun phrases in the three texts. As mentioned earlier, this constraint posits that it is unlikely to have two lexical arguments in one clause core. This constraint was, in general, supported by the three texts, but with variation among them. FD shows the most support with no incidence of two lexical NPs within a single clause core; MSA comes in middle with 3 incidences, and CA shows the least support with 9 incidences. Table 3 below summarizes the results under this constraint in the number of incidences and the total percentage.

Table 3. Results for Constraint 1 "avoid more than one lexical argument".

	Number of Lexical arguments per clause core			
	0	1	2	Total
FD	77 (68%)	36 (32%)	0 (0%)	113
MSA	73 (50%)	70 (48%)	3 (2%)	146
CA	61 (44%)	69 (45%)	9 (6 %)	139

Constraint 2. Avoid Lexical A.

The second constraint focuses specifically on the A role. When there is one lexical argument, the three varieties generally agreed to avoid putting lexical arguments in the A slot, with some differences among them. FD data has only 2% of the lexical arguments in the A position and 77% in the O position. MSA has 13% of the lexical arguments in the A position and 64% in the O position. CA has 18% of the lexical arguments in the A position and 59% in the O position. Tables 4, 5 and 6 below summarize the results of the second constraint in each data source.

Table 4. The number of lexical arguments per clause core in CA data.

	Forms				
Role		Lexical	Pronominal	Ø	Total
	A	14 (18%)	21 (28%)	41 (54%)	76 (36 %)
	O	49 (66%)	23 (31%)	3 (4%)	75 (35 %)
	S	15 (37.5%)	15 (37.5%)	10 (25%)	40 (19 %)
	Sn	15 (68%)	6 (28%)	0	21 (10 %)
	Total	93 (43.5%)	65 (31%)	54 (25%)	212

Table 5. The number of lexical arguments per clause core in MSA data.

	Forms				
Role		Lexical	Pronominal	Ø	Total
	A	6 (13%)	6 (13%)	33 (73%)	45 (24%)
	O	33 (73%)	11 (24%)	1 (2%)	45 (24%)
	S	16 (26%)	16 (26%)	30 (48%)	62 (33%)
	Sn	25 (66%)	13 (34%)	0	38 (20%)
	Total	80 (42%)	46 (24%)	64 (34%)	190

Table 6. The number of lexical arguments per clause core in FD data.

Table 6: The number of lexical arguments per clause core in FD data.					
	Forms				
Role		Lexical	Pronominal	Ø	Total
	A	1 (2%)	19 (53%)	16 (44%)	36 (25%)
	O	27 (77%)	7 (20)	1 (3%)	35 (24%)

	Forms				
	S	8 (12.5%)	35 (55.5%)	20 (32%)	63 (43%)
	Sn	1 (8%)	11 (92%)	0	12 (8%)
	Total	37 (26%)	72 (49%)	37 (25%)	146

Recall that constraints 1 and 2 are *grammatical* (presented above), and constraints 3 and 4 are *pragmatics* that deals with the activation cost of the core arguments (to be presented below). The results also supported the two pragmatic requirements with variations among the three data sources.

Constraint 3. Avoid more than one new argument per clause.

The third constraint states that only one core argument per clause core carries new information. The three data sources supported this constraint. Table 7 below shows that in both MSA and FD data, no clauses were found that contained more than one *new* argument. This should come as no surprise since clauses with two lexical arguments were limited anyway in the texts of these two varieties, as the results of constraint 1 revealed. The CA data showed only 2% of clauses that contained two new arguments in a single clause core. The word order seems to play a role here, as discussed later in the discussion section. Overall, the information pressure (or the newness) within the clause core was low in the three data sources, as shown in Table 7.

Table 7. Results for Constraint 3 "avoid more than one new argument".

	Number of new arguments per clause core			
	0	1	2	Total
FD	98 (87%)	15 (13%)	0	113
MSA	106 (73%)	40 (27%)	0	146
CA	76 (55%)	60 (43%)	3 (2%)	139

Constraint 4. Avoid new A.

The last constraint states that if we have one core argument that carries new information, then the speaker should avoid encoding the new argument in the A slot. This constraint was also supported by the data. Table 8 shows that the distribution of new arguments in CA data was as follows: only 9% from the A role were lexical new NPs (i.e., 9% heavy A argument, 46% heavy O argument, 30 % heavy S argument, and 48 % heavy Sn argument). The MSA data supported constraint 4 even more, as shown in Table 9. In MSA data, only 5% were heavy A arguments in comparison to 44% of heavy O, 14.5% of heavy S role, and 24% of heavy Sn arguments. Finally, recall that the FD data exhibited only 2% lexical arguments in the A role (see constraint 2). Therefore, FD data did not have any new argument in the A role in comparison to 23 % of heavy O and 9.5 % of heavy S role, and 8% of heavy Sn. The results of FD are summarized in Table 10 below. Note the overall *newness* or information pressure in FD data was minimal across all arguments. The low-pressure information in FD indicates implications regarding the number of audiences involved, the familiarity between speaker and hearer and the shared cognitive awareness of the topic between them, and the genre of the text itself (i.e., non-expository speech). Some of these implications will be addressed in the discussion section.

Table 8. Arguments by information status and syntactic role in CA data

Role	Information Status			
		Given	Accessible	New
	Total			
A	64 (84%)	3 (4%)	9 (12%)	76(36%)
O	35 (46%)	5 (7%)	35 (46%)	75(35%)

	Information Status				
	S	27 (67.5%)	1 (2.5%)	12 (30%)	40(19%)
	Sn	8 (38%)	3 (14%)	10 (48%)	21 (10%)
	Total	134 (63%)	12 (6%)	66 (31%)	212

Table 9. Arguments by information status and syntactic role in MSA text

	Information Status				
Role		Given	Accessible	New	Total
	A	42 (93%)	1 (2%)	2 (5%)	45 (24%)
	O	22 (49%)	3 (7%)	20 (44%)	45 (24%)
	S	51 (82%)	2 (3%)	9 (14.5%)	62 (33%)
	Sn	26 (68%)	3 (8)	9 (24%)	38 (20%)
	Total	141 (74%)	9 (5%)	40 (21%)	190

Table 10. Arguments by information status and syntactic role in FD text

	Information Status				
Role		Given	Accessible	New	Total
	A	36 (100%)	0	0	36 (25%)
	O	23 (66%)	4 (11%)	8 (23%)	35 (24%)
	S	55 (87%)	2 (3%)	6 (9.5%)	63 (43%)
	Sn	11 (92%)	0	1 (8%)	12 (8%)
	Total	125 (86%)	6 (4%)	15 (10%)	146

To sum up, the quantitative results of the four constraints confirm PAS reality in the data. However, the three varieties differ among them in holding consistency to those constraints. FD exhibited the most consistency, followed by MSA, followed by CA. The next section is devoted to a general discussion of these constraints in Arabic with examples from the data.

7. Discussion

The data show that the clause core was mostly light, and the argument distribution resembles what has been reported about PAS in different languages. Consider the following examples of the three varieties:

CA clauses

(1) qad ʔašraṭ-a l-ḏaalika nafs-a-h w aʔawbaqa
 PRTCL obligated-3msg to-that self-ACC-3msg and weakened.3msg

diin-a-hu li-ḥuṭaami-n yantahzu-hu ʔaw maqnabi-n yaquudu-hu
 faith-ACC-3msg for-wreckage-INDEF win.3msg -it or prey-INDEF drag.3msg-it

ʔaw minbari-n yafraṣu-hu qad ʔaaman-a min šaxši-hi
 or platform-INDEF claim.3msg-it PRTCL bend-3msg from body-3msg

w qaarab-a min xaṭwi-hi wa šammar-a min ḥaubi-hi ...
 and bring-closer-3msgb from steps-3msg and roll-3msg from robe-3msg

‘He obligated himself to the war, and weakened his faith to win some wreckage, to drag some prey, to claim a public platform. He bent himself (in preparation), brought his steps closer, and rolled his robe.’

MSA clauses

(2) ?u-stuqbil-tu fii kull makaan fii l-ʕaalam l-ʕarabii
PASS-welcome-1msg in all place in the-world the-Arabic
‘I was welcomed in all places in the Arabic world.’

(3) ?aʕaad-u li-ya ?aθ-θiqa-ta bi-nafsii
brought-back-3pl to-me the-confidence-fsg in-myself
‘They brought me back my confidence in myself.’

(4) ?aʕaad-u li-ya ?aθ-θiqat-a fii ?albahθ
brought-back-3pl to-me the-confidence-fsg in the-search
‘They brought me back the confidence in the search.’

FD clauses

(5) ?aqbal-t haddaa-t ?im-diir-a gaʕad-d θamma
turned-1msg came-down-1msg the-hometown-f stayed-1msg there

?arbaʕa-t ?ayyaami-n w ?awjah-t
four-f days-INDEF and returned-back-1msg
‘I turned and went down to the hometown, stayed there for four days, then returned Back.’

(6) ?il-mihim haaš-a wa-llah nafaʕ-a-nii w dawwar-a l-ii
the-important went-3msg by-God help-3msg-1sg and searched-3msg to-3sg

li-malaabsi-n
to-cloths-INDEF
‘The important thing, he went and helped me, he searched for some clothes for me.’

These examples and many others like them in the data exhibit consistency with the four constraints of PAS. The clauses, above, contain transitive verbs that describe events in which an agent performs an action affecting a patient. The agent in these clauses is mainly a human and is known to the listener from the beginning of the speech context. Take for example the following clauses: [ʔašraʔa] ‘**he** obligated X’; [aʔawbaqa] ‘**he** weakened X’; [yantahzu] ‘**he** wants to take X’; [ʔaʕaadu] ‘**they** brought back X’; [nafaʕa] ‘**he** helped X’. The speakers of these clauses whether CA, MSA, or FD, shared similar cognitive assumptions about their addressees at the time of articulation. The three speakers began by establishing the human referents at the beginning of each segment. Then, they anticipated the activation process in their audience’s mind and made the agents of subsequent transitive verbs *given* as A role. This is predictable according to PAS constraints because the semantic focus in the transitive verbs is not the agent but rather the patient or the complement. The O role in the examples above was preferred to carry the heavy information such as the lexical nouns and the new material.

Pragmatics played an important role in the management of the information in the discourse. Although the three speeches were all spontaneous (presumably CA too as the written text describes it), they seem to vary according to the size of the

audience. As mentioned earlier in the description of the texts (Section 5.1), FD data involve only one-person audience since the data was taken from a recorded conversation between two friends, MSA data involve around fifteen people audience since it was an interview in a live studio, and one can just assume that CA data involve a larger size of audience than FD and MSA since it was (khutbah) performed in front of gathered crowd, as it was described in the written form. With the audience size in mind, we find the clause core of CA and MSA exhibited more lexical NPs and new information than FD. Consider the following examples from the three speeches which show how each speaker encoded lexical NPs and new information in the subject role:

CA clauses

(7) kalam-u-kum yuhii aṣ-ṣumma aṣ-ṣilaab
talk-NOM-3mpl weakens.3msg the-rocks the-strong
'Your talk weakens the strongest rocks.'

(8) fiṣl-u-kum yuṭmiṣu fii-kum ṣaduwwa-kum
action-NOM-3mpl covet.3msg in-you enemy-your
'Your action covets your enemy.'

(9) baqiya rijaal-u-n ġaḏḏa ʔabṣaar-a-hum ðikr-u
remain men-NOM-INDEF lowered.3m gaze-ACC-3mpl remembering-NOM

ʔal-marjiṣ
the-resurrect
'The memory of the resurrect day lowered their gazes,'

(10) w ʔaraaq-a dumuuṣ-a-hum xaufu ʔal-maḥṣar
and drop-3m tears-ACC-3mpl fear the-judgment-day
'And their tears drop fearing from the judgment day.'

MSA clauses

(11) ʔal-jumhuur ʔas-suurii ʔistaqbal-a tajaarib-ii
the-fans the-Syrian welcome-3msg experiences-POSS
'The Syrian fans welcomed my experiences.'

(12) haḡa-l-jiiil huwa man yaḥmil janiin l-mustaqbal
This-the-generation they COMP carry.3msg fetal the-future
'This generation carries the fetal of the future.'

(13) ʔas-salaam qad yaxluq ʔaṭ-ṭamaʔniina
the-peace PRTCL creates.3msg the- tranquility
'Peace might create tranquility.'

FD clauses

(14) ġiṣṣa-t il-maṭaar tiṣajjib

story-fsg the-airport laugh.fsg

'The airport story makes someone laugh.'

(15) yanšur nikilas yabsit hinii ʔaha w zimalaay-uu

go-at-night Nicolas hangout.3msg here he and peer-3msg

'Nicolas comes at night to hangout here with his peers.'

The proportion of heavy clauses (containing lexical NPs and new information) varies among the three data, and it seems to be linked to the size of the audience. More audience leads to more heavy clauses, and this can be seen across the three data. FD discourse involves only a single addressee, and it exhibited the least information pressure among the three data. FD data did not involve any transitive verbs with two full lexical arguments (see Tables 6 and 10 in the quantitative results section). MSA discourse involves approximately fifteen people, and it exhibited more heavy clauses than FD (see Tables 5 and 9). CA data exhibited more heavy clauses than FD and MSA (see Tables 4 and 8). We do not know the size of the audience in the CA speech, but it can be speculated from the book's description of the speech (Khutbah) that the audience number is big. Therefore, it seems that there is a correlation between the information pressure in discourse and the size of the audience involved, but such a correlation requires further investigation and confirmation.

Now, let us turn back to the set of examples above (sentences 7-15). The subjects in all these clauses were new and lexical NPs. In CA, the subject NPs were [kalamukum] 'your words', [fiʕlukum] 'your action', [ðikru ʔal-marjiʕ] 'the thought of judgment day', and [xaufu ʔal-mahšar] 'the fear of judgment day'. In MSA, the subject NPs were [ʔal-jumhuur ʔas-suurii] 'the Syrian fans', [haθa l-jiiil] 'this generation', and [ʔas-salaam] 'the peace'. Finally, in FD, the subject NPs were [gišsat il-maṭaar] 'the airport story', and [nikilas] 'Nicolas' (person name). Let us close examine these subject NPs.

To begin with FD, when the speaker realized that the storyline requires a new lexical NP, he avoided placing it in the A slot and chose, instead, to place it in the S slot. PAS constraints posit that the S slot is open for realizing lexical NPs or new information. It is the A slot that dislikes new and lexical information. The FD speaker could have easily said [gišsat il-maṭaar ʕajjabannii] (AVO) 'the airport story made me laugh', but he chose to say [gišsat il-maṭaar tiʕajjib] (SV). He could have also said [yanšur nikilas hinii] [AVO] 'Nicolas came here', but he chose to say [yanšur nikilas] [SV]. Notice that even though these verbs offer two argument slots (i.e., transitive form), the speaker chose the intransitive form to convey the new information as lexical NPs. This confirms PAS constraints and suggests that the information pressure of lexical SV is lighter than lexical AVO.

The information pressure in MSA speech is heavier than in FD. Consider the clause from 11 above [ʔal-jumhuur ʔas-suurii ʔistaqbala tajaaribii] [AVO] 'the Syrian fans apprehended my experiences'. Here, the two slots welcomed the two different lexical NPs that carry new information. This clause and, similarly, the ones in 12 and 13 are considered the heaviest clauses in MSA data because the subject and the object of the transitive verb are filled with lexical NPs and carry new information in each clause. PAS constraints posit that these clauses are rare in spontaneous speech. MSA has only 2% of such heavy clauses.

However, the heavy clauses of MSA can be justified if we consider the so-called *rhythmic principle* of MSA (Hole, 2004). This principle suggests that the two heavy arguments within one clause core are often separated by the verb, which means that heavy clauses appear as SVCOMP (subject-verb-complement) instead of VSCOMP. The rhythmic principle lightens a heavy clause by introducing the heavy elements on the edges of the clause with the verb in the middle as a break with a voice pause sometimes (i.e., avoiding the adjunct position of heavy elements).

Finally, CA speech included heavy clauses than MSA and FA. Clauses 7-10, above, are examples of lexical and new NPs in both A and O roles. Nevertheless, the heavy clauses in CA make up a small proportion of the data even though we cannot be sure whether the spontaneous speech was edited and to what extent. Still, it is remarkable to see PAS constraints are manifested even in the speeches prior to 800AD.

With this support of PAS in spoken Arabic, the question for future empirical work remains: would PAS constraints hold

as well in the Arabic written materials? It has been suggested that PAS constraints were visibly overcome in the writing system of other languages (see Mazur-Palandre, 2015; and Lin 2009).

8. Conclusion

This study has shown that PAS is a characteristic of Arabic spontaneous speech. The discourse analysis of the three data sources has shown that PAS is manifested in different types of Arabic spontaneous speech whether Classical Arabic, Modern Standard Arabic, and the spoken dialect (Faifi Arabic in the present study).

The observed patterns in the data indicate the following key points regarding the processing of information in Arabic spontaneous speech:

1. The A argument is generally disfavored for encoding heavy information such as lexical NPs and new information.
2. When heavy information occurs in a clause core, the O argument is strongly preferred to carry the heavy information.
3. Despite the difference in speech genre, the era, and the number of audiences involved in a speech, the speakers create a similar systematic distribution of referents especially encoding information in A and O argument slots.

While further research is needed to determine the generalizability of these key points about Arabic discourse, it is hoped that this study has shed some light on our understanding of the links between grammar and discourse. Future research should examine new areas such as planned vs. unplanned speech events, planned vs. unplanned written material, multiple speech productions by the same speaker, and L1 vs. L2 acquisition. Such avenues underscore the potential wealth of information still to be uncovered about the topic.

REFERENCES

- Allen, S. (2000). *A discourse-pragmatic explanation for argument representation in child Inuktitut*. *Linguistics*, 38, 483–521.
- Allen, S. (2007). Interacting pragmatic influences on children's argument realization. In M. Bowerman & P. Brown (eds), *Crosslinguistic perspectives on argument structure: implications for learnability*, 191-210. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Allen, S., & Schröder, H. (2003). Preferred argument structure in early Inuktitut spontaneous speech data. In J. Du Bois, L. Kumpf, & W. Ashby (Eds.), *Preferred argument structure: Grammar as architecture for function*, 301-338. Amsterdam: John Benjamin's Pub.
- Ashby, W., & Bentivoglio, P. (2003). Preferred argument structure across time and space: A comparative Diachronic analysis of French and Spanish. In J. Du Bois, L. Kumpf, & W. Ashby (Eds.), *Preferred argument structure: Grammar as architecture for function*, 61-80. Amsterdam: John Benjamin's Pub.
- Ashby, W., & Bentivoglio, P. (1993). Preferred argument structure in spoken French and Spanish. *Language Variation and Change*, 5(1), 61-76.
- Bloom, P. (1990). Subjectless sentences in child language. *Linguistic Inquiry*, 21, 491-504.
- Brickell, T. C., & Schnell, S. (2017). Do grammatical relations reflect information status? Reassessing Preferred Argument Structure theory against discourse data from Tondano. *Linguistic Typology*, 21(1), 177-208.
- Chafe, W. (1994). *Discourse, consciousness, and time: The flow and displacement of conscious experience in speaking and writing*. Chicago: University of Chicago Press.
- Chui, K. (1992). Preferred argument structure for discourse understanding. *Proceeding of COLING-29*, 1142-1146. Retrieved October 29, 2015, from <http://www.aclweb.org/anthology/C92-4180>
- Clancy, P. (2003). The lexicon in interaction. In J. Du Bois, L. Kumpf, & W. Ashby (Eds.), *Preferred argument structure: Grammar as architecture for function*, 81-108. Amsterdam: John Benjamin's Pub.
- Du Bois, W. (2003). Argument structure. In J. Du Bois, L. Kumpf, & W. Ashby (Eds.), *Preferred argument*

- structure: *Grammar as architecture for function*, 11-60. Amsterdam: John Benjamin's Pub.
- Du Bois, W. (1987). The discourse basis of ergativity. *Language* 63, 805-855.
- Fillmore, C. (1984). Pragmatically controlled zero anaphora. *Twelfth Annual Meeting of the Berkeley Linguistics Society*, 12, 95-107.
- Everett, C. (2009). A reconsideration of the motivations for preferred argument structure. *Studies in Language*, 33(1), 1-24.
- Genetii, C., & Crain, L. (2003). Beyond preferred argument structure: Sentences, pronouns, and given referents in Nepali. In J. Du Bios, L. Kumpf, & W. Ashby (Eds.), *Preferred argument structure: Grammar as architecture for function*, 197-223. Amsterdam: John Benjamin's Pub.
- Holes, C. (2004). *Modern Arabic: Structures, functions, and varieties*. Georgetown University Press.
- Jiang, X., Zhang, F., Yan, R., & Chen, L. (2022). Preferred argument structure in the oral narratives of adolescents with and without SLI. *Clinical Linguistics & Phonetics*, 1-17.
- Huang, C. (2012). Preferred argument structure in mandarin child language. *Taiwan Journal of Linguistics*, 10(2), 119-168.
- Huang, H., & Huang, S. (2009). Beyond preferred argument structure the discourse pragmatics of noun phrases in Tsou. *Studies in Language*, 33(3), 499-527.
- Kohn, S., & Cragolino, A. (2003). The role of preferred argument structure for understanding aphasic sentence planning. In J. Du Bios, L. Kumpf, & W. Ashby (Eds.), *Preferred argument structure: Grammar as architecture for function*, 339-351. Amsterdam: John Benjamin's Pub.
- Kumpf, L. (2003). Genre and preferred argument structure: Sources of argument structure in classroom discourse. In J. Du Bios, L. Kumpf, & W. Ashby (Eds.), *Preferred argument structure: Grammar as architecture for function*, 109-130. Amsterdam: John Benjamin's Pub.
- Lin W. (2009). Preferred argument structure in Chinese: A comparison among conversations, narratives and written texts. *North American Conference on Chinese Linguistics*, 2, 341-357.
- Matsumoto, K. (2000). Intonation units, clauses and preferred argument structure in conversational Japanese. *Language Sciences*, 22(1), 63-86.
- Mazur-Palandre, A. (2015). Overcoming preferred argument structure in written French: Development, modality, text type. *Written Language & Literacy*, 18(1), 25-55. Chicago
- Mittelberg, I., Farmer, A., & Waugh, R. (2007). They actually said that? An introduction to working with usage data through discourse and corpus analysis. In Gonzalez-Marquez, M. (Ed.), *Methods in cognitive linguistics*, 18, 19-52. John Benjamins Publishing.
- Owens, J., Young, W., Rockwood, T., Mehall, D., & Dodsworth, R. (2010). Explaining Ø and overt subjects in spoken Arabic. *Information structure in spoken Arabic*, 20-60.
- Prince, E. (1981). Toward a taxonomy of given/new information. In P. Cole, (Ed.) *Radical pragmatics*, 223-255, New York: Academic Press.
- Serratrice, L. (2002). *Syntax and pragmatics in the acquisition of Italian subjects*. Paper presented at the *Ninth International Congress for the Study of Child Language*, Madison, WI.
- Weber, E. (2003). Nominal information flow in the talk of two boys with autism. In J. Du Bios, L. Kumpf, & W. Ashby (Eds.), *Preferred argument structure: Grammar as architecture for function*, 354-383. Amsterdam: John Benjamin's Pub.