Dorota Molin*

Intransitive clause word order in Neo-Aramaic: information structure, pragmatics and word order shifts

https://doi.org/10.1515/flin-2024-2051
Received January 10, 2023; accepted September 10, 2024; published online November 4, 2024

Abstract: This paper studies quantitatively intransitive constructions in North-eastern Neo-Aramaic, and their implications for Neo-Aramaic word order typology (historically, 'subject-verb-object'). Though not considered previously, transitivity proves to be a significant factor in Neo-Aramaic word order variation. The differences between the intransitive Subject and the transitive Agent in Neo-Aramaic are a product of their divergent information-structural tendencies, as well as, it would seem, their basic syntactic preferences. Unidentifiable (i.e. indefinite, old) Subjects largely follow the verb, which fits with the larger tendency for pre-verbal arguments to be topical. Identifiable arguments, however, have roughly equal pre- and post-verbal frequencies in the corpus, when jointly considered. A deeper analysis considers referential distance and lexical-pragmatic factors. I propose that VS for identifiable Subjects is more likely for functions associated with 'discourse discontinuity', being less likely for foreground events. This distribution in Neo-Aramaic hints at a functional versatility of the VS structure, which is thus not restricted to a single construction such as sentence focus. By contrast, in (co-territorial Qəltu) Arabic, transitivity and topicality are less significant for word order variation. Finally, comparative data from Neo-Aramaic varieties of different word order profiles reveal the instability of the VS clause amidst larger word order changes. The cross-dialectal data show that the loss of the VS clause correlates very closely with the 'verb-object' to 'object-verb' shift. While Aramaic would have started as '(Agent-)verb-object', some Neo-Aramaic dialects show increasing rates of 'object-verb'.

Keywords: word order; transitivity; information structure; discourse pragmatics; Aramaic; Arabic

***Corresponding author: Dorota Molin**, University of Oxford, Oxford, UK,
E-mail: dorota.molin@ames.ox.ac.uk. https://orcid.org/0000-0002-2407-2012

1 Introduction

1.1 Word order variation: a corpus-based typology with a focus on discourse

This study investigates the intransitive clause in Northeastern Neo-Aramaic ('NENA', 'Neo-Aramaic') in the context of information structure, discourse pragmatics and general word order tendencies in NENA.[1] Northeastern Neo-Aramaic is a group of contemporary vernacular Aramaic descended from the Northwestern branch of Semitic, whose many varieties are now critically endangered. NENA has historically been spoken by Jewish and Christian communities living in adjacent regions of Southeastern Turkey, Northern Iraq, and Western and Northwestern Iran (see further Section 1.2).[2]

This Neo-Aramaic subgroup is often described as 'SVO' in the literature.[3] This word order tendency is most likely inherited from pre-modern Aramaic, whose relatively recent attestations usually tend to subject-verb(-object), though often with a (considerable) degree of variation.[4] Consequently, scholars consider subject-verb-object to be the most conservative linearization of transitive word order in Northeastern Neo-Aramaic,[5] though some dialects have now developed other orders in the transitive clause (see below).

---

1 Glottocode nort3241. On the corpora, see further Section 1.4. below.

2 I thank the three anonymous reviewers, whose comments helped improve the article significantly. All errors and omissions remain mine.

3 For transitive clause typology in NENA, see Noorlander and Molin (2022), which also includes a brief section on the intransitive subject. The paper by Noorlander et al. (forthcoming) compares subject-verb-object and subject-object-verb varieties of NENA, showing that the latter is most likely an innovation.

4 For an overview of Late Antique Aramaic, see Kutscher (1952: 126). For earlier forms, see e.g. Kaufman (1974: 133, 160), Folmer (1995: 521–587); Gzella (2015: 177, 180). While none of the written forms of pre-modern Eastern Aramaic constitute a direct antecedent of Northeastern Neo-Aramaic, the general prevalence of the subject-object-verb tendency in the historical varieties points to such a configuration as the starting point also for Northeastern Neo-Aramaic. This is also expected on comparative Semitic grounds. This notwithstanding, OV orderings did feature prominently in several varieties of ancient Eastern Aramaic (specifically 'Imperial Aramaic'), often appearing alongside VO. For an overview, see e.g. Folmer (1995: 521–587). The aforementioned historical studies do not resolve/address the issue of markedness, nor do they study different object types in detail; some scholars (e.g. Kutscher 1952: 126) claim that the word order is 'free'.

5 E.g. Khan 2009 (344–345); Noorlander forthcoming; Noorlander and Molin 2022; Noorlander et al. forthcoming.

Still, such accounts of Northeastern Neo-Aramaic as 'SVO' omit the intransitive subject, or fuse it with its transitive counterpart.[6] This, in turn, leaves unanswered questions about the distinctions or similarities between these two subject types. Moreover, the statistical dominance of intransitive clauses in many types of discourse (cf. Haig and Schnell 2016) means that the intransitive domain is key to our understanding of a language's syntax (Dryer 1997; 2007: 79).

The general semantic and discourse-functional differences between transitive and intransitive clauses are well known. To begin with, the intransitive subject role is semantically more versatile than that of the transitive one. In many languages, the former covers agents, experiencers and patients (e.g. Keenan 1976; Comrie 1988: e.g. 111; Andrews 2007; Dryer 2007; Haig et al. 2022; Haspelmath 2011; Keenan 1976). These differences have motivated scholars such as Andrews (2007) to propose a split of the 'subject' (spelled here with lowercase 's') into two grammatical roles: the transitive 'A(gent)' and the intransitive 'S(ubject)'. This model is taken up in the present paper and applied to word order, thus paralleling the studies by Du Bois (e.g. 1987b), Haig and Schnell (2016), Bickel (e.g. 2008, 2011) or Donohue (e.g. 2008).[7] Consequently, in this paper we use the terms AVO and AOV orders for transitive clauses and SV/VS orders for intransitive clauses.

At the same time, the implications of these insights into the differences between the Agent and the Subject are still to be properly investigated for the discourse and syntax of many languages. The present study investigates the effects of 'topicality' and non-topicality[8] on word order linearizations in Neo-Aramaic, and to a smaller degree, in Neo-Arabic of the same region.[9] The question of argument typology and corpus distribution is highly relevant for Aramaic, which – like other Semitic languages – has a highly pragmatically-sensitive word order (e.g. Hoberman 1989; Khan 2008: 855–80). Still, this pragmatic sensitivity has not yet been systematically correlated with specific argument types.

6 See, for instance, Khan's grammar of the Christian Barwar NENA variety (NW Iraq; 2008: 855–864).

7 Some word order studies incorporate transitivity as a core variable, even while using the label 'S' for both types of subject; see e.g. Siewierska (1988; 1997b), Siewierska and Bakker (2009: 296–299), Jacennik and Dryer (1992) and Dryer (2007, e.g. 17). The 'S' versus 'A' label is especially common in typological approaches to alignment; see e.g. Dixon (1979, 1994), Comrie (1978: 118), Bickel (e.g. 2011).

8 For these concepts, see e.g. Prince (1981), Chafe (1987), Lambrecht (1994, 2000), and Lambrecht and Polinsky (1997).

9 Aside from word order, there are also a number of (recent) studies on semantics/topicality and argument expression (null, pronoun or lexical). For instance, the pro-drop tendency of the Agent has traditionally been correlated with its high topicality and referential continuity (e.g. Givón 1983, Du Bois 1987b), and, more recently, with persistence. However, some authors (e.g. Bickel et al. 2015) caution against strong universalist generalizations about the referential hierarchy.

Indeed, the close investigation of word order in NENA in Section 3. below challenges the notion of a single, unified 'subject' category, at least as far as word order is concerned. The dialects representative of conservative word order (i.e. AVO) show roughly equal rates of post- and pre-verbal Subjects when no other variables are considered. This, in turn, justifies their separate treatment in the present quantitative study, which allows us to identify the specific variables interacting with syntactic variation. This paper considers the distribution of intransitive subjects, based for the most part on Jewish varieties of Neo-Aramaic from northwestern Iraq. In the corpus studied, 52 % of all intransitive arguments are pre-verbal (96 out of 184 nominals and pronominals). By contrast, the rate of pre-verbal As in all clauses is 94 % (12/190).[10] The difference between the two subject types is also significant if we consider only nominal arguments, as shown in Table 1 below.[11]

Table 1: Positional frequencies of lexical Ss and As.

	XV% (*n*)	VX% (*n*)	Total *n*
A	91 % (99)	9 % (10)	109
S	49 % (81)	51 % (83)	164
Total *n*	180	93	273

($p < 0.05$).

In accounting for the 'verb-subject' clause in Neo-Aramaic, this paper draws from the notion of sentence focus and theticity, which have often been studied in the context of languages tending to AVO/SV.[12] Sentence-focal clauses are defined in terms of their *information structure*: they do not contain any 'presupposed' (i.e. old, topical or, typically, definite) referents, and are thus described as 'all-new' (e.g. Lambrecht 2000). Thetic statements are similar, but are usually defined in terms of their *pragmatic* function and *cognitive* structure: they are deictic in nature, drawing attention to a situation and presenting it as 'an undivided whole' (Sasse 1987; see also Matras and Jürgen-Sasse 1995). Both structures thus contrast with categorical statements, which name an entity ('the topic') and make a statement about it ('comment'), and so have a bi-partite structure. Thetic and sentence-focal constructions are used, for

10 As regards argument indexing on the verb, this is obligatory for transitive and intransitive subjects in most paradigms, in the four NENA types represented here. The indexing of objects, by contrast, is differential.

11 See Section 1.4. for the methodology.

12 For NENA, a very brief outline of 'sentence focus' may be found in Coghill's study of information structure in the Christian dialect of Telkepe (2018: 319–325). Both the Telkepe variety and conservative dialects at the heart of this study share an AVO profile (Sections 2–3 below), including the general AVO profile and the existence of VS.

instance, in answers to the question 'What happened?'.[13] Since sentence focal and thetic constructions make use of 'subject inversion', it has often been claimed that this inversion encodes the subject as a non-topic (Lambrecht 2000: 618–619).[14] This makes good pragmatic sense in AVO languages, since the prototypical slot of the 'topic' is pre-verbal, the topic being prototypically the Agent. At the same time, the 'verb-subject' order for non-topics is also documented for Latin, whose 'classical' word order is AOV (Bolkestein 1995). In this study, I draw from Lambrecht's approach to sentence focus, but employ this term in a broader, functional sense. I use it for *all* predications with all-new information status, rather than only for those constructions which express their all-new status with a marked structure, as he does (Lambrecht 2000: 617).

Another function of the 'verb-subject' clause, relevant to the Neo-Aramaic variation considered below and closely related to sentence focus and theticity, is the presentative. In this case, the *new* situation presented is specifically the appearance of a new discourse referent. The predicate-argument structure in presentative constructions is documented for, *inter alia*, Germanic, Slavic and Romance. It is attested even in the syntactically rigid English, which makes use of the *there* inversion (e.g. *There came a man*).[15] We should emphasize here already, however, that the Neo-Aramaic 'verb-subject' clause is not restricted to indefinite referents.

Generative studies have argued that the post-verbal slot is 'default' with the so-called unaccusative subjects, which tend to correspond with the 'non-agentive' class in functional typology.[16] Even this phenomenon, however, is arguably often linked to definiteness,[17] rather than (only) to subject semantics. For the Neo-Aramaic study below, definiteness – and especially its lack – is also shown to be a better predictor of variation than subject semantics (e.g. non-agentivity).[18]

---

13 Thus, according to Lambrecht, 'sentence–focal' and 'thetic' (cf. Sasse 1987, 516–520) share their information structure, but not necessarily formal (syntactic or phonological) features (Lambrecht 2000: 619).

14 Lambrecht (2000) does not relate 'sentence' focus to the issue of object position explicitly, beyond writing that VS occurs in 'AV(O) languages'.

15 Its structurally near-identical counterparts in Dutch and German would be respectively, *Er kwam een man* and *Es kam ein Mann*.

16 See Levin and Rappaport Hovav (1992) for a semantic approach to 'unaccusativity' within a generative framework. See also Tortora (1999) for Italian specifically; both deal with movement verbs specifically.

17 See Section 1.4. below for the definition.

18 Compare the Italian *Si è rotto un vaso,* 'A VASE broke' (indefinite S; sentence-focal interpretation), and *La mia tazza si è rotta,* 'My cup broke' (definite; topic-comment structure). For Italian, see e.g. Arnaiz (1997, 50).

1.2 The broader typological, historical and comparative value of Neo-Aramaic word order

The significance of such a corpus-based typology for Neo-Aramaic is not only language-internal. The well-documented microvariation in NENA along with its various contact-based parameters, and the unparalleled diachronic depth of Aramaic can help produce fine-grained models for various large-scale and cross-linguistically-attested tendencies (see e.g. Gutman 2018: 1). This motivates including here a small comparative excursus in Section 4. Though brief, this section identifies variables and proposes a model of word order change that can serve as the starting point for larger, comparative and diachronic studies. This diachronic-comparative survey contextualizes the intransitive clause within the general word order profile of several Neo-Aramaic varieties, and compares it with Neo-Arabic (Jewish Qəltu of northern Iraq).

The comparative NENA data show word order tendencies across several different clause types, and across four typologically different NENA varieties which show increasingly advanced stages of the AVO > AOV shift. This synchronic variation can be correlated with different stages of diachronic processes, in line with the historical-comparative method (see e.g. Harris and Campbell 1995, 345–359; Haig 2008).[19] It appears that the maintenance of the VS clause in Neo-Aramaic is closely dependent on the (statistical?) prevalence of (A)VO in the transitive domain, or at least on VO with discourse-new objects. A future study should compare the situation in Neo-Aramaic with that of ancient Aramaic, where OV also occurs, and in some corpora even with non-focalized and non-topical objects. The variation of objects versus subjects, however, does not seem to exhibit the same kind of relationship as the one suggested here for NENA.[20]

As regards the comparison of Neo-Aramaic with Neo-Arabic, it reveals that some Arabic varieties (still) make wider use of the VS clause. This difference supports the diagnosis that in Neo-Aramaic – in a partial distinction to Arabic – the positional difference between A and S is not only due to their distinct discourse functions, but also due to the variable of transitivity itself.

1.3 Neo-Aramaic and its dominant orders

The Northeastern Neo-Aramaic varieties spoken today belong to one of the last surviving branches of Aramaic. Aramaic, in turn, is a sub-group of the North-West

19 This method is especially useful for languages whose direct ancestors are not attested, as is the case with NENA, but is applied even for richly-documented language families.

20 See, for instance, the work on Biblical Aramaic by Owens and Dodsworth (2009; on subjects) or Cook (1986; on objects).

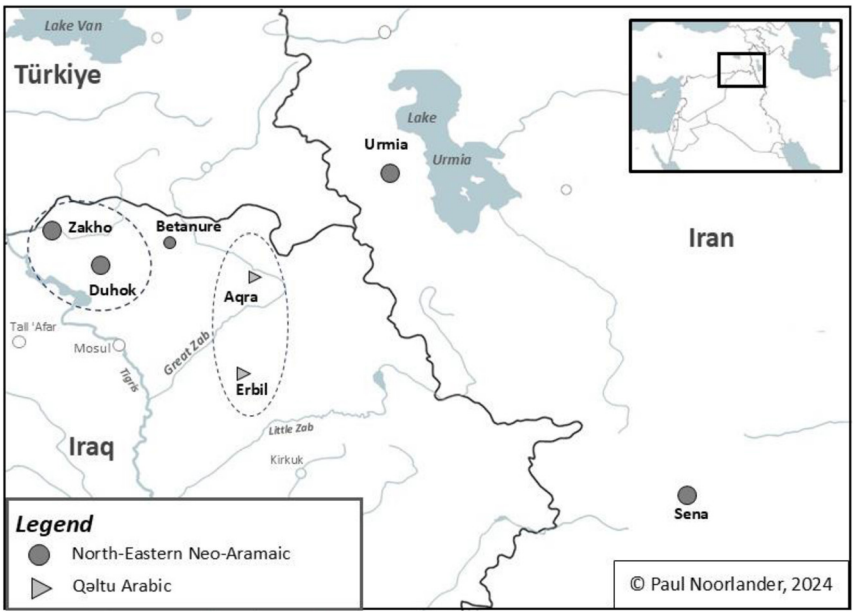

Figure 1: The North-Eastern Neo-Aramaic area, with the Aramaic and Arabic doculects studied indicated.

Semitic group, boasting some 3,000 years of documented language history. North-eastern Neo-Aramaic is a highly diversified group that consists of different languages,[21] spoken (historically) by Jewish and Christian communities in south-eastern Turkey, northern Iraq and western Iran (e.g. Khan 2007, 3–4). The specific Neo-Aramaic and Neo-Arabic varieties studied in this paper are shown on the map in Figure 1 and in Table 2 below. Living for centuries in relative socio-geographical isolation within a largely mountainous region, these communities have preserved a language group that was widespread throughout the Near East, before the diffusion of Arabic in the region (e.g. Gzella 2015).

For centuries, NENA and its ancestor language have lived in the context of intensive language contact with languages from several families, especially with Northwest Iranian languages (known today as Kurdish and Gorani), with which it now shares a territory (e.g. Haig and Khan 2019; throughout). This history of intensive contact has impacted NENA in both lexicon and grammar, including word order.

21 In this paper, both 'dialects' and 'varieties' are used to refer to specific NENA types for the sake of stylistic variety. The fact is that the differences between the NENA varieties amount to cross-linguistic, rather than cross-dialectal, distinctions.

Table 2: Corpus metadata.

Provenance	Name in the article	Glottocode	Communal affiliation	Dominant transitive word order	Data source	Number of informants
Duhok and Zakho (NW Iraq)	AVO Neo-Aramaic	lish1247	Jewish	AVO	Duhok: fieldwork by Molin; Zakho: Meehan and Alon (1979)[a]	5
Betanure (NW Iraq)	Neo-Aramaic tending towards AVO	lish1247	Jewish	AVO but definite Os shifting preverbally	Mutzafi (2008a)	2
Urmia (NW Iran)	Neo-Aramaic tending towards AOV	urmi1250	Christian	AOV with relics of VO	Khan (2016), vol. 4 (henceforth '2016 $_{IV}$')	>10
Sena/Sanandaj (W Iran)	AOV Neo-Aramaic	sana1291	Jewish	AOV	Khan (2009), excl. poem E[b]	4
Erbil/Arbel and Aqra	J. Qəltu Arabic	qelt1235	Jewish	AVO	Jastrow (1990)	5 (most likely)

[a]The total corpus length was around 7,500 words. For the number of the relevant clauses analyzed in the other doculects, see Table 10 in Section 4. [b]Some of the annotations and counts for the Christian Urmia and Jewish Sena NENA varieties were carried out by Paul Noorlander and are due to appear in Noorlander (forthcoming), as indicated in Section 3 below.

Significantly for this paper, several dialects of NENA spoken on its eastern periphery (see Figure 1 above) have now shifted from the inherited Aramaic AVO order to AOV,[22] converging with the pattern in the neighbouring Iranian. Indeed, the existence of other contact-induced features and the high volume of loanwords in this group of dialects (e.g. Garbell 1965; Khan 2009; Mutzafi 2008b; Noorlander 2014) corroborate the idea that AOV too is a contact-induced innovation. This also fits the social situation; the speakers of AOV Neo-Aramaic live in relative isolation from the Neo-Aramaic heartland, and often in especially intense situations of contact with Northwestern Iranian languages (e.g. Khan and Mohammadirad 2024). The shift

22 'Dominant' word order is defined here as the statistically dominant order within the corpus studied (cf. Siewierska 1997b). These 'dominant' orders in NENA also overlap with the 'basic' word order, as defined by Siewierska (1988: 8). Note also that the label 'A' (as opposed to 'S') is preferred in this paper, in order to isolate the transitive subject from its intransitive counterpart. Thus, tags such as 'SOV' are replaced here with 'AOV'.

from VO to OV has been explored on a cross-section of NENA in a number of recent (corpus-based) studies.[23] In the present paper, this OV profile is represented by the dialect of the Jews of Sena in western Iran (henceforth 'AOV Neo-Aramaic'),[24] and of the Christians of Urmia in Iran's North-West.[25] As will be shown, however, the typology of the latter remains somewhat mixed, which in diachronic terms implies that the word order changes are less advanced, and likely still ongoing. In contrast, the dialect of the Jews of Duhok and Zakho in northwestern Iraq represents the more conservative AVO typology.[26] Finally, the Jewish dialect of Betanure (also northwestern Iraq; Mutzafi 2008a) is in general very close to the Duhok and Zakho varieties, both linguistically and communally. Still, the Betanure variety shows slightly higher rates of OV (for definite arguments), alongside diminishing rates of VS. As I suggest in Section 4.2., these two phenomena seem to be related.

1.4 Methodology: doculects, corpora and variation

Important to this paper's view of argument structure is the notion of grammatical functions, as mentioned above and as articulated by Andrews (2007). In this model, the analysis of argument structure is language-specific. The general categories of grammatical relations (subject, object) can be further subdivided if its members exhibit distinct behaviour and/or coding properties (Andrews 2007: 3–4). This paper distinguishes between the intransitive argument 'S' and the transitive subject 'A', which are tagged primarily on the basis of their behaviour and coding properties, and only secondarily through semantic roles such as 'agent'.[27] For objects, the general grammatical relation 'object' is sufficient for this paper's purposes, and subsumes both the object of monotransitive ('Patient') and the often affectee-like object of ditransitive ('Theme', or 'direct object' in Aramaic studies).[28]

The core study in Sections 2–3 is based on the Duhok/Zakho Neo-Aramaic variety, with three additional groups of Neo-Aramaic and one group of Neo-Arabic appearing in Section 4. These varieties are shown in Table 2 below. The corpora used comprise

23 See especially Haig and Khan (2019); Khan (2009); Noorlander and Molin (2022); Noorlander (forthcoming), and Noorlander et al. (forthcoming).

24 For a grammar, see Khan (2009).

25 For a grammar, see Khan (2016), vols. 1–2.

26 For a grammar, see Molin (2024).

27 See e.g. Comrie (1988) and Dixon (1994). For NENA argument structure from a similar perspective to mine, see Noorlander and Molin (2022) and Noorlander (2021: 58–60).

28 These two object types exhibit a similar syntactic behaviour in NENA (cf. Noorlander and Molin 2022). For the morphosyntax of objects in NENA, see further Coghill (2014; on differential object marking). For morphosyntactic alignment in Neo-Aramaic, see Coghill (2016; diachrony), and Noorlander (2021; synchronic typology).

narratives with past deixis (often with recurring discourse topics) and – to a lesser extent – descriptions of customs and traditions.[29]

The core study of this paper (Sections 2–3) concentrates on intransitive clauses in Jewish Duhok/Zakho dialects of the conservative AVO profile. For this survey, I tagged the below-listed features of argument and information structure, in intransitive clauses with nominal (i.e. lexical) Subjects. Independent pronominal subjects were also counted in both transitive and intransitive clauses, but were not tagged for the following features.

1. Goals/sources:
   'Goals/sources' refers to both destinations and sources of movement verbs. Across NENA (as well as the neighbouring Iranian languages), goals tend to be clause-final; in NENA, the same generally holds for sources.[30] Furthermore, in NENA and Neo-Arabic, goals/sources tend to be prepositionally marked, but this is optional in some of the doculects studied here (e.g. Jewish NENA of Sena and J. Qəltu Arabic). This means that in the corpora, goals/sources appear mostly as oblique, but also as direct arguments.
   1.1. The presence of a goal/source with intransitive movement verbs;

(1) *u* *awa* *yɪʔl-le* ***l-ʔoya***|
and he enter.PRT-3ms inside
'And he came inside.'
(Own fieldwork)

   1.2. The absence of a goal/source
2. Agentivity:[31]
   2.1. Agentive referents: humans, agentively-behaving animals and divine beings;

(2) ***t͡ʃut͡ʃɪky-aθa*** ***là*** *aθe* *axl-i-lu.*|
bird-PL NEG come.3PL eat-A.3PL-O.3PL
'…[so that] birds wouldn't come and eat them.'
(Own fieldwork)

29 For statistical significance tests, Chi Square (total data count over 20) or Fisher Exact test (total data count under 20) were used see (Levshina 2015: 213–214).

30 See the volume edited by Haig et al. (2023), where more detailed, corpus-based data on goals can be found for a series of languages across Western Asia. For NENA specifically, see Noorlander and Molin (2022).

31 Other semantic features such as 'control' (Croft 2012) were not tagged systematically but were still considered in relevant cases (see further Section 1.2. above and Section 2.2. below). In reality, 'control' is a matter of a semantic spectrum, as recognized by Croft himself. Thus, 'humanness' is preferred here as a more basic and semantically discrete feature.

2.2. Non-agentive referents: inanimate entities;

(3) *āθe-wa* *ʾeð-ɪt* *zyàra.*|
come.3ms-PST.IPFV feast.ms-GEN visiting
'[Then] the feast of Pentecost would come.'
(Own fieldwork)

3. Identifiability (i.e. definiteness; cf. e.g. Lambrecht 1994: 90–93)
This tag approximates the category of semantic definiteness. Since definiteness marking is not fully grammaticalized (i.e. obligatory) in AVO Neo-Aramaic, 'identifiability' was assigned purely on semantic-referential grounds, based on each individual context.

3.1. Identifiable
Specific and identifiable referents explicitly mentioned in previous discourse, implicit therefrom or made indexically identifiable (e.g. with a possessive index);[32] generally-known and/or unique referents.[33] Generic referents (e.g. 'wheat') were also tagged if they were anaphoric (see example), since in such cases, it is the broader class (e.g. 'wheat') which is being treated as topical in the discourse.[34] Overall, 'definite' referents are thus also 'given', perhaps except for anaphorical generics.

(4) *gòr-a*| *g-ɪb-e* *-wā-la* *raba* ***bàxt-e.***|
husband-ms REAL-love-A.3ms -PST.IPFV-O.3fs much wife.fs-POSS.3ms
'A [certain] husband loved his wife very much.'
(Own fieldwork)

'Wife' is not mentioned previously, but implicit from the mention of 'a certain husband'.

3.2. Unidentifiable
Referents not present – explicitly or implicitly – in the preceding discourse, even though they may be specific; mass/group entities whose specific referents is unknown (e.g. 'people', but not 'the people of that town'; see Lambrecht 1994: 90–93).

---

32 For the full list of implicit reference types, see e.g. Jacennik and Dryer (1992: 213).

33 This variable largely corresponds to the 'specific' tag in the study of Arabic by Owens et al. (2009, 48–49). However, 'specific' also includes referents that, while concrete, have not yet been made identifiable in the discourse, that is, are semantically indefinite. An example would be the subject 'student' in 'A student approached me to discuss a problem with me after the lecture.' See also the study on Brazilian Portuguese by Naro and Votre (1999), who use the tags 'evoked'(i.e. definite in some way) and 'available'. However, they appear to presuppose that 'availability' is a feature of lexical semantics, rather than discourse.

34 Definite coding of anaphoric generics is reported for English by Lambrecht (1994: 90–93), and is evidenced in NENA by their optional differential object marking (Molin 2024: 138), and by their definite marking with demonstratives.

(5) *ɪθ-wa* *xa* ***nàš-a***| *tˤālɪb-wa* *nedavòt.*|
EXIST-PST INDEF man-ms ask.3ms-PST.IPFV alms
'There was a certain man who used to beg for alms.'
Beginning of a story
(Own fieldwork)

4. Referential continuity, activation
This variable serves the specific function of tracking the 'activeness' of a given referent in discourse, and thus adds a layer of complexity to the simple binary of 'identifiable' versus 'nonidentifiable' (Variable 3). Variable 4. serves to capture the increasing degrees of mental accessibility of a referent (e.g. Givón 1983: 9–11, Chafe 1987).
4.1. Inactive
Referents not previously activated in discourse. This includes referents in principle inferable from the previous context, but not yet explicitly signalled as a discourse topic. This is illustrated by the referent *baxte* 'his wife' in example (4), implicit by the mention of 'a husband'.
4.2. Active referents (part of 'activated')
Referents active in the immediately preceding discourse (up to four clauses back);[35]

(6) Previous clause: 'The people gathered for the Sabbath prayers.'
ur-ra *šàbθa…*ˈ
enter.PRT-3fs Sabbath.fs
'The Sabbath began…'
(Own fieldwork)

4.3. Semi-active referents (part of 'activated'): active 5–15 clauses 'ago', but not in the immediately preceding discourse.

Other variables whose statistical effects could hardly be tested because of their infrequency, were not tagged. This includes argument weight (three or more phonological words per argument), backgrounding/subordination and polarity. In other corpus studies of Neo-Semitic, these variables were tagged, but often turned out to be weak/statistically non-significant.[36]

35 These may thus include 'chain initial topics' (Givón 1983: 9), as well as frequently overlap with 'aboutness topics', i.e. the topic of the clause in question (cf. Lambrecht 1994: 117–127; Givón 1983: 9).
36 For weight in Arabic, see Holes (1995), Hewitt (2006) and Al-Shawashreh (2016: e.g. 75, 97–98). In Al-Shawashreh's quantitative study, weight appears to be non-significant. For the issue of backgrounding and clausal context, see, for instance, Dahlgren (1998), but the variable 'backgrounding' involves some speculation, in particular in the absence of overt encoding of clause types, as is often the case in Neo-Aramaic.

This study utilises the methods of corpus-based, quantitative typology, which recognizes that for many features, typology is a matter of gradience. As such, the variation should be captured with a set of (quantitative) variables, and modelled on a spectrum (see e.g. Haig et al. 2022). I follow the method of descriptive statistics for corpus linguistics developed by Levshina (2015). The Chi square is employed to test significance, as long as the total number of observations is above 20 and none of the expected frequencies are below 5. Otherwise, the Fisher exact test is used (Levshina 2015: 213–4). For all tests, the value of degrees of freedom is 1, unless stated otherwise, and the threshold of significance is 0.05.

2 The general word order trends (AVO Neo-Aramaic)

This sections gives an overview of the basic differences between the S, A and O arguments in AVO Neo-Aramaic in terms of semanto-pragmatics and referential density. It also presents the variables which are relevant for the distribution of the S, and shows data on two variables which are excluded from the core discussion. The statistically significant factors interacting with the distribution of Subjects are: the presence of Goals and identifiability. The strongest association is found between unidentifiable and non-activated referents, and the VS order. The semantic parameter of humanness has no direct relevance; referential continuity also interacts less directly with word order (see further Section 3 below).

2.1 Basic pragmatic and semantic differences between the S, A and O

The distinctive features of the Neo-Aramaic S – including the way it contrast with the A – are partly anticipated by the following data on the density of lexical referential expression and identifiability.[37] First, intransitive subjects show significantly higher rates of lexical expression than their transitive counterpart (Table 3):[38]

37 For the density of lexical expression across arguments, see for instance, Haig et al. (2022: 145–157). Haig et al. report relatively stable rates of subject anaphors, though they do not distinguish transitivity. For a perspective that takes into account the NP's morphosyntax (i.e. case) for the rates of lexical expression, see Bickel (2003). Cross-linguistic data about the density of lexical expression per argument may be found at MULTICAST; https://multicast.aspra.uni-bamberg.de/.

38 This count was made on the basis of a large subsection of the corpus, containing around 80 % of the total S tokens.

Table 3: Lexical expression for transitive versus intransitive subjects.

	NP	Zero	Total *n*
Intransitive subject	20.5 % (110)	79.5 % (426)	536
Transitive subject	7 % (75)	93 % (960)	1,035
Total *n*	185	1,386	1,571

($p < 0.01$).

As is the case cross-linguistically, this difference is an effect of the high topicality rates among the A arguments. As identifiable referents, most Agents function as discourse topics and are thus not explicitly referenced with a nominal (or an independent pronoun).[39] By contrast, the higher rate of lexical S implies that the intransitive subject covers a wider range of functions in discourse, including non-topics and types of topics which are more likely to have a lexical argument (see Section 3 below). A similar tendency is shown by identifiability. In the corpus of AVO Neo-Aramaic, 97 % of nominal Agents are identifiable (106/109),[40] though sometimes 'only' by implication. The rates of definite Subjects are significantly lower in the corpus, though still high: 69 % (112/162). They imply that the Subject is also used for non-topical arguments, which is further studied below.[41]

2.2 The 'verb-medial' tendency in the transitive and intransitive domain

The first major factor influencing the distribution of Subjects in AVO Neo-Aramaic considered here is the presence of a goal/source of movement ('G'). In the intransitive clause, the Subject and the goal tend to be placed on different sides of the predicate, rather than crowded on one side of the verb. Combined with a trend to place goals/sources clause-finally in NENA, this results in a dominance of the SVG order in clauses with both a Subject and a goal, as shown in Table 4 below. However, a larger corpus would be necessary to corroborate the dominance of this order. This common SVG order is given in (7a) below, while the less common VSG is illustrated in (7b). Both clauses in (7) have an unidentifiable subject, most of which occur in the post-verbal

39 See the classical paper by Du Bois (1987b) with its 'avoid lexical A' constraint, and the response thereto by Haig and Schnell (2016). See also Bickel (2003) and Bickel et al. (2015), and Haig et al. (2022).

40 This is consistent with the cross-linguistic data in Haig et al. (2022: 164), where rates of 'givenness' – as well as humanness – were always above 90 %.

41 I lack the data on the identifiability of objects, though it is likely not (drastically) lower than that of the Subject. Haig et al. (2022: 164) report the cross-linguistic mean for 'givenness' of objects of 75 %, though with significant variation in the mean of the language corpora included.

Table 4: Frequencies of S positions with and without goals/sources.

	SVG[a]	VSG/VGS	Total *n*
Goal present	76 % (19)	24 % (6)[b]	25
No goal present	44 % (60)	56 % (77)	137
Total *n*	79	83	162

($p < 0.05$). [a]No SGV or GSV clauses were recorded. [b]Of this: 5 cases of VSG, 1 of VGS.

slot (see Section 3.1. below). The presence of a goal/source seems to override this tendency, as shown in (7a) below.

(7) (a) SVG
rix-a *là* *basim-a* *k-eθe* *mɪn* *pɪmm-e.*|
smell-ms neg good-ms REAL-come.3ms from mouth-POSS.3ms[42]
'[Some] bad smell is coming out of his mouth.'
(Own fieldwork)

(b) VSG
u-k *-eθe-wa* *ràba* *naʃ-e* *kɪs* *bab-i…*|
and-REAL -come.3ms-PST.IPFV many man-PL at father-POSS.1cs
And many people [from the neighbourhood] used to come to my father [for treatment].'
(Own fieldwork)

Given the limited amount of data, the motivation for this tendency to place the S on the opposite side of the goal/source remains somewhat uncertain. Still, this distribution mirrors the ordering preference in the transitive clause. There, 98.5 % (69/70) of clauses with both a nominal A and an object place their predicate *between* these two arguments. When all clauses with an A are counted regardless of the presence of the object, by contrast, the Agent is somewhat more flexible in its position, appearing preverbally in 91 % of cases.[43] Both the transitive and the intransitive clauses, therefore, strongly prefer the order subject-verb-argument (or subject-verb-object in intransitive clauses) in clauses with (at least) two nominal arguments (constituents), and seem to allow for greater subject flexibility if the additional argument is missing. This could be a general syntactic, 'verb-medial', preference, perhaps akin to 'crowding' avoidance. This has been observed for verb-second languages (e.g. German),[44] and some AOV systems, where, however, it seems motivated by different

42 Nominals are tagged for person-number-gender inflection only if they agree with the verb.
43 See Table 1 Section 1.1. above.
44 See, for instance, Lambrecht (1994: 194).

reasons.[45] Because of their effect on Subjects, clauses with goals/sources are thus excluded from the following analysis or set apart therein, in order to maximize the visibility of other variables.

The aforementioned verb-medial trend suggests that goals/sources of intransitive verbs behave syntactically like an argument, namely like the object of the transitive clause. This could turn out to be semantically motivated, which, in turn, could be investigated by comparing the distribution of goals to that of other adjuncts.[46] Briefly, the available corpus does not reveal such a verb-medial preference for locatives, temporal or manner adverbials when all are counted jointly, though their Total *n* are few. With these types of adjuncts, both VS and SV clauses occur in the corpus. The example in (8a) below gives a clause in which the Subject and the adjunct are 'crowded together' on one side of the predicate, while in (8b), these two constituents are split by the predicate. It seems that specific adjunct types have their own positional preferences, such as the clause-initial placement of 'scene setting' adverbials (see (8a) below). A more detailed investigation of such adverbials is yet to be carried out.

(8) (a) XYVS clause[47]
qabɪl tre-alpe u-tre-ɪmma ʃɪnn-e,| axxa k-ħakm-ì-wa|
before two-thousand and-two-hundred year-PL here REAL-rule-3PL-PST.IPFV
yawnǎn.|
Greece
'Two thousand two hundred years ago, Greece used to rule here.'
(Own fieldwork)

(b) XSV clause
xà-yarxa,|[48] *palg-ɪt èn-a pθɪx-le.|*
one-month half.ms-GEN eye-POSS.3fs open.PRT-3ms
'[After] one month, half of her eye opened.'
(Own fieldwork)

45 Ueno and Polinsky (2009) find that the frequency of two-argument predicates in AOV languages is lower, relatively to AVO. They interpret this as a pressure to avoid the appearance of more than one argument before the verb (ibid). However, the cognitive motivation for the crowding avoidance possibly remains distinct in the two cases. Ueno and Polinsky argue that in AOV, the dispreference may be due to a processing difficulty of predicate-final utterances (2009: 479). This is not relevant for AVO or SV/VS structures; for SVG, see further Section 3.2. below.

46 One semantic factor may be the integration of goals (and often sources) within the State of Affairs in contrast to adjuncts. Contrast the centrality of the phrase 'in/to the shop' in sentences 'I *went to* the shop' versus 'I was shopping *in the shop*'. For more on movement verb semantics and 'unaccusativity', see also Levin and Rappaport Hovav (1992) and Tortora (1999).

47 'X' and 'Y' indicate here adjuncts.

48 As this example shows, prepositional marking can be optional in some cases.

2.3 Humanness: no direct significance

Aside from goals, another variable which is largely excluded from the core study in Section 3. is humanness. In this case, this is because Subject humanness is not significant ($p > 0.05$); the results are summarized in Table 5 below. The results are very similar when clauses with goals are included (i.e. the observed and expected frequencies are very similar). Moreover, even among clauses with human Subjects, the Subjects of highly agentive predicates such as 'to run' or 'to go out'[49] appear both before and after the predicate. This point is relevant since Lambrecht (2000) seems to assume that sentence focus – at least when it is expressed with the VS order – correlates with non-agentive events. By contrast, the acceptability of VS with agentive predicates in NENA problematises an account of the VS clause in terms of a single function. It also calls for a more precise definition of 'sentence focus'.

3 Discourse parameters: a deeper look at identifiable and unidentifiable referents

The Subject variation in NENA correlates with pragmatic and partly syntactic factors. The pre-verbal position tends to be restricted to definite arguments. This means that pre-verbal arguments (Subjects, but also others) are therefore interpreted as conveying some sort of topicality. This results in a strong tendency to place unidentifiable subjects post-verbally. Identifiable subjects, by contrast, are common in both positions, implying that VS cannot be reduced to sentence-focal constructions. At the same time, the post-verbal placement of definite referents is more likely in specific clause types, and I suggest that it tends to correlate with specific discourse functions, subsumed here under discourse discontinuity. Unlike in some other languages, word order variation in NENA is thus not directly caused by a

Table 5: Positional frequency of Ss by animacy (without goal clauses).

	SV	VS	Total *n*
Human	42 % (34)	58 % (46)	80
Inanimate	46 % (26)	54 % (31)	57
Total *n*	60	77	137

49 These verbs are at the top of Croft's (2012: 257–258) hierarchy of agent-like subjects.

semantically-motivated alignment split, but rather due to information-structural and discourse-based tendencies.[50]

First, word order variation in conservative NENA closely interacts with discourse parameters. Thus, identifiability is a significant variable ($p < 0.05$) of moderate explanatory power. It is particularly unidentifiable referents which show an especially strong correlation with word order: unidentifiable referents appear very often in the post-verbal position, as shown in Table 6 below. The reverse, however, does not apply – identifiable referents only slightly prefer SV.[51] When goal/source clauses are included, the VS rate for unidentifiables drops somewhat – to 74 % (37/40). Both groups of Subjects – identifiable and otherwise – are investigated further in this section.

3.1 Unidentifiable referents: the discourse functions of VS

The typical post-verbal position for new information is shown in the sentence in (9a) below.[52] The first discourse function associated with such VS clauses is that of 'theticity' or 'sentence focus'. This is illustrated in (9a), which appears at the beginning of a story. The sentence thus answers the implicit question 'What happened' to

Table 6: Positional frequency of identifiable and unidentifiable subjects (goal clauses excluded).

	SV	VS	Total *n*
Unidentifiable	17 % (7)	83 % (33)	40
Identifiable	55 % (53)	45 % (44)	97
Total *n*	60	77	137

($p < 0.05$).

50 Word order ergativity has been proposed for 'a number of languages', including Parï (Nilotic, South Sudan; Dryer 2007; citing Andersen 1988. In Parï the transitive OVA (in the article, 'OVS') contrasts with the intransitive SV. Even in Parï, however, the pre-posing of Ss is apparently not semantically restricted, but rather linked to information structure (cf. Andersen 1988). Semantic ergativity is also proposed for Palu'e – an Austronesian language from South Indonesia. Palu'e is characterized as AVO (Donohue 2008: 37–38), but non-agentive Ss and objects (mostly) follow the verb. I am not aware of how this maps on information structure.

51 This distribution is thus somewhat opposite to languages such as Palu'e (Austronesian; AVO). In the latter, the restriction affects the 'opposite' position, from the point of view of NENA. According to Donohue (2008: 37–38), all subjects in Palu'e can precede the verb, but only non-agentive ones can follow it.

52 This has also been found for the Christian Telkepe, where – according to Coghill – subject as a discourse-new referent 'appears to always come after the verb' (2018: 322–323).

the man introduced at the outset of the story, and has a non-topical subject (cf. Lambrecht 2000: 618–619).[53] At the same time, the few 'thetic-like' clauses in SV (illustrated in (9b)) demonstrate that VS is not obligatory even in such event-oriented assertions. Additionally, as already mentioned in Section 2.2. and illustrated in the example in (7a) above, the presence of a goal/source also makes pre-verbal subject placement more likely.[54]

(9) (a) All-new status; VS
xa-yoma *o* *naʃa* *mpɪq-la* *gulɪk-ta* *go* *kàs-e*…|
one-day DEM.ms man go_out-PRT.3fs ball-fs in belly-POSS.3ms
'One day this man – a swelling appeared in his belly…'
(Own fieldwork)

(b) All-new status; SV
xa-yoma *xa-mɪqubal* *mɪn* *huzay-e* *mət-le.*
one-day INDEF-Qabbalist.ms from Jews-PL die.PRT-3ms
'One day, a certain Qabbalist from [among] the Jews passed away.'
(Meehan and Alon 1979: XV: §1)

The post-posing of newly-introduced referents in event-oriented assertions is manifested especially strongly in clauses with the verb *payɪʃ* (*p-y-ʃ*) 'to break out, occur', whose S is virtually always postverbal. Clauses with *payɪʃ* are inchoative existential predications used with non-human referents, such as atmospheric phenomena or events, e.g. 'a war (broke out)', 'a storm (broke out)'. In such assertions, the VS order is so strongly grammaticalized that it is used even with identifiable entities. Thus, in (10) below, 'the Farhud Pogrom' has already been mentioned in the narrative (10 clauses ago), and can thus be considered semi-active:

(10) Context: events leading to the Farhud – the pogrom on the Jews of Baghdad, 1941.
pɪʃ-le *farhù̯d*…|
appear.PRT-3ms Farhud
'(Then, finally,) the Farhud pogrom broke out (…and they killed people).'
(Own fieldwork)

The second discourse function associated with post-verbal Subjects occurs in clauses that bring new referents into the discourse. This is illustrated in (11) below, where the character is introduced with the verb 'to come'. This lexical choice is likely related to

53 It should be noted that the example in (9a) does include one topical element, namely 'this man', but this referent is extra-clausal, i.e. 'detached' or 'extraposed' (I thank the reviewer for pointing this out). In general, the influence of extraposed elements on word order has not been studied yet.
54 As reported in Table 6 above, 7/40 of unidentifiable Subjects appear before the subject. In clauses with Goals, the rate of new Subjects before the predicate is 6/10.

the conceptualisation of the introduction of characters as their appearance 'on the stage', as further explored below (Section 3.3.).[55]

(11) Context: 'My husband was in the hospital, the Mount Scopus hospital [and I was with him].'
b-lèle| θe-le dòxtor,| θe-le xă-nåsax| mɪn-levanòn.|
in-night come.PRT-3ms doctor.ms come.PRT-3ms INDEF-sick.ms from-Lebanon
At night, A DOCTOR came, [and also] a patient from Lebanon came.'
(Own fieldwork)

A comparable tendency to place new characters after the predicate also holds for existential clauses, indicating that it is a wider pragmatic tendency, applying across clause types. In conservative NENA varieties, such clauses employ the 'predicate-argument' order by default.[56] The predicate is an uninflected existential particle (e.g. *iθ, hit*), as illustrated in (12) below:

(12) *ɪθ-wa xa-beθ-a* (…) *d-ʕāyɪʃ-wa mɪn sʿìwe.|*
EXIST-PST INDEF-house-ms SBR-live.3ms-PST from wood
'There was a house[hold] which used to make its living from wood.'
(Own fieldwork)

3.2 Identifiable referents: a general positional freedom

Identifiable referents occur in both positions, as illustrated in (13a–b) below. This near-minimal pair in (12a–b) demonstrates that the two permutations are possible under (almost) identical pragmatic and semantic configurations: both have an inanimate but definite subject. There are, however, factors making VS more likely, as discussed in the following section (3.3.).[57]

(13) (a) *aql-e tʿrʿìs-la.|*
leg.fs-POSS.3ms heal.PRT-3fs
'His leg healed!'

55 Several of such clauses, in turn, are also compatible with the sentence-focal interpretation. An example thereof may be viewed in (9b) above, where the deceased Qabbalist is introduced in the very sentence that announces his death.

56 The inverse, pre-predicate position is only possible for arguments in narrow (e.g. counter-presuppositional) focus, like in the sentence 'At school, we learnt to read, but there wasn't *writing (*in NENA, 'writing – there wasn't).'

57 Further examples with highly-referential, discourse–topical referents are given in Section 3.3. below.

(b) *baθr arbi yom-e tˁrˁìs-le reš-i.*|
after forty day-PL heal.PRT-3ms head.ms-POSS.1SG
'In the course of forty days, my head healed.'

These findings are compatible with those for spoken (e.g. Peninsular) Arabic, where 'definite, specific nouns' are said to 'escape a unitary profile' (Owens et al. 2009: 54),[58] only offering a few further observations. The following section, by contrast, shows that for Aramaic, further variables relevant to word order variation can be identified. Indeed, there are other AV(O) languages where the functional versatility of the VS seems relatively broad.[59] For instance, in a corpus of written, formal Polish narrative, only 35 % (18) cases of VS are explained as 'presentative' (with new referents).[60]

3.3 Identifiable referents in VS: discourse discontinuity

The remaining question is, therefore, whether identifiable referents are completely free, or whether we can identify further variables correlating with their position. In this section, I show that what goes a considerable way towards explaining their distribution are specific verbal lexemes. This, in turn, is due to the discourse functions of the verbal lexemes in question, rather than being a direct product of the verbs' lexical semantics.

First, while all intransitive verb types are attested in the VS structure, the verbs 'to come' and 'to rise' do so significantly more often. This trend is most clearly visible for the verb 'to come', which is relatively more frequently represented in the corpus, and appears with both identifiable and unidentifiable referents.[61] This verb's distinctive distribution among all other intransitive verbs is highly significant, both

58 Likely, however, the freedom of definites is even greater in many varieties of Arabic (including Kurdistani) than it is in AVO NENA. See Owens et al. (2009) and Section 4.1. below.

59 For Polish, see Jacennik and Dryer (1992). In that paper, the (presentative) VS(/A) clause is distinguished from VS(/A)X, which in fact is shown to host topical subjects even more frequently than the SV clause. These counts do not distinguish A and S.

60 In the spoken NENA corpus, the frequency is 45 % (cf. Table 6 above). In the Polish corpus, for instance, VS is also said to occur for the function of 'participant removal', which accounts for 12 % of VS (Jacennik and Dryer 1992: 228–229). As regards transitivity, 73 % of VS/A (but not VS/AX) clauses in the Polish corpus are intransitive; the (corresponding) figure in the NENA corpus is even higher (89 %).

61 Most other motion and manner of motion verbs could not be studied statistically due to their low overall frequencies, though they may also perform similar functions (e.g. 'to enter' may function comparably to 'to come' in some cases). Also, the total count of SV clauses with 'to come' would have been higher had clauses with goals not been eliminated, since 'to come' co-occurs with goals rather frequently. Still, what the data in Table 8 show is that without goals, the argument of 'to come' is almost always post-verbal.

Table 7: The distribution of 'to come' versus all other verbs (no goals).

	SV	VS	Total *n*
'To come'	14 % (5)	86 % (31)	36
Other verbs	54.5 % (55)	45.5 % (46)	101
Total *n*	60	77	137

($p < 0.01$).

in clauses with and without goals. Table 7 reports the data for clauses without goals. The findings for the verb 'to rise' point in a similar direction, but the tokens are too low at present for a robust hypothesis: in 9/9 cases, its subject is also post-verbal.

This difference in distribution is even more striking when only identifiable referents are considered (again, no goals for control). The rates of 'to come' remain similar to those reported in Table 7, but all other verbs shift towards SV (Table 8).

A closer look suggests that such VS clauses with 'to come' and 'to rise' tend to correlate with structural 'seams' in the discourse. They activate implicit topics, reactivate existing ones, create narrative cohesion and signal new narrative units.[62] Thus, they do not tend to convey narrative foreground events. First, it is known that the verb 'to come' is used to introduce new referents. This function is found cross-linguistically (Du Bois 1987a), but is restricted to those referent introductions which can be construed as a spatial 'arrival' onto the scene of events (Schnell et al. 2021).[63] This has already been shown for Neo-Aramaic in (11) above, where the introduction of 'a doctor' into discourse coincides with the doctor's arrival – likely to the narrator's hospital ward. But in the Neo-Aramaic corpus, 'to come' is also common with identifiable referents:

(14) (a) Context: 'We were sitting together, thinking what to do. We felt hungry and sad.'

θe-le	*xatɪn*	*xàθ-i*
come.PRT-3ms	husband.ms.GEN	sister-POSS.1cs

'My brother-in-law came (and said "let's dance").'

Table 8: The distribution of 'to come' versus other verbs, with identifiable referents (no goals).

	SV	VS	Total
'To come'	16 % (3)	84 % (16)	19
Other verbs	64 % (50)	36 % (28)	78
Total	53	44	97

($p < 0.05$).

62 This function has also been identified for Christian Telkepe NENA (Coghill 2018: 321).

63 For example, this verb is thus unexpected for introducing inanimate referents.

(b) Context: Food processing in the home
k-ɪθy-a-wa *yɪmm-i* *g-maql-à* *-wa-lu.*|
REAL-come-3fs-PST.IPFV mother-my REAL-clean-A.3fs -PST.IPFV-O.3PL
'…my mother would come and clean them (i.e. the animal parts which we had bought).'
(Own fieldwork)

It seems that in these cases, 'to come' with unidentifiable and identifiable referents fulfils a similar role, which is supported by further observations about the clauses in (14). First, both clauses lack a source/goal, and arguably even a (*n* obvious) need for their referents' change of location. As regards the protagonist in (14a), he is already implicitly present at the scene. The previous discourse describes the extended family sitting together, discussing what to do. So 'my brother-in-law's' change of location is minimal, suggesting that another function of 'to come' might also be involved. More specifically, while the referent 'my brother-in-law' has so far been implicit, in the clause in (14a), he is flagged as an active discourse topic; indeed, this is the first instance when he actively contributes to the events.

The function of the clause in (14b) is akin to topic activation: it *specifies* a previously implicit topic. The speaker's mother is clearly involved in the preceding discourse, which addresses the conservation and storing of food in the speaker's family. But the preceding clauses use first person plural forms: 'We used to by legs, heads, and stomachs [of sheep].' In (14b), however, the mother is singled out in order to spell out the duty performed specifically by her: the cleaning of animal parts.

Such clauses with 'to come', therefore, serve to cast the spotlight onto a different topic, while being construed as the character's arrival at the scene of events. 'To come' thus offers in such clauses a useful alternative to the existential clause. As is the case cross-linguistically (Abbott 1997), in Neo-Aramaic, the existential appears to be dispreferred for referent presentation when the referent is definite. Thus, examples like (15), which rephrases (14a) with the existential, are unattested in the Duhok–Zakho corpora:

(15) ***ɪθ-wa* *xatɪn* *xàθ-i*
EXIST.PST husband.ms.GEN sister-POSS.1SG
'There was my brother-in-law.'

This function explored for the examples in (14) is applicable to many other clauses with 'to come'. For instance, with animate (i.e. potentially topical and mobile), identifiable but inactive referents which also lack a goal, the subject is always postverbal (9/9). This corroborates the idea that the VS in such clauses correlates with topic switch, activation or specification.

Importantly, the use of VS with such 'to come' clauses in the corpus of AVO Neo-Aramaic is possible with all levels of topic activation (see Section 1.4. above), rather than being dependent on a significant distance from the character's previous mention. Thus, clauses with 'to come' which fit especially well the function of 'topic reactivation' described above (no goal, animate) occur with inactive referents (9/9 in VS), with semi-active ones (2/2 in VS),[64] but also with active ones (6/9 in VS). In other words, the signalling of re-presentation with VS (with the verb 'to come') is apparently an optional stylistic device available to a speaker, rather than a strict tool for mental retrieval of semi/inactive-active referents.[65]

At the outset of this section, I noted that the verb 'to rise' also displays a strong preference for VS with identifiable subjects (9/9), and this too can be explained with recourse to discourse function. 'To come' in Neo-Aramaic has a discourse-inchoative and discourse-cohesive function, which respectively refer to new units of discourse and causal binding to the preceding.[66] Like with 'to come', however, most of the clauses are also compatible with a physical interpretation (in this case, rising). For instance, the example in (16) below represents a new narrative unit, which matches the fact that this event takes place on a separate day:

(16) Context: 'The man and the wolf spoke in the forest. The wolf went back to his place, the man went home [and they went to sleep].'
durdɪt-yom q-qayɪm mărе bèθa,| g-ezɪl ta sʿìwe.|
second-day REAL-arise.3ms master.ms.GEN house REAL-go.3ms for wood
'The next day, the master of the house gets up [again] and goes for wood.'
(Own fieldwork)

This section has studied clauses which reactivate, switch or make explicit referents which are identifiable, and which open new narrative sections. I focused specifically on the verbs 'to come' and 'to rise', which are used for such discourse functions and are represented relatively richly in the corpus. When occurring with identifiable referents, these two verbs show significantly higher rates of VS, compared with all other intransitive predicates. Crucially, therefore, clauses with topic switches or reactivation, and with changes in narrative units resemble clauses with new referents: all show a strong preference for post-verbal Subjects. We can now therefore propose a meta-category that unites these different instances of VS: discourse

64 These counts exclude clauses with goals to increase the likelihood that they correlate with topic re-introduction or switch (see above in this section).

65 The latter is apparently more relevant to the issue of zero (/pronoun) versus or nominal expression of a topic. In several languages, lexically expressed topics have been correlated with an increased referential distance from their previous mention (see the edited volume by Givón 1983).

66 For 'cohesion', see Cohen (2012: 290–293), who describes the verb 'to rise' as presenting a protagonist's reaction to a preceding event (like 'therefore').

discontinuity.[67] 'Drawing from Givón (1983), 'discourse discontinuity' refers here to 'seams' in the narrative structure; a new direction in the flow of the discourse with regard to the active ('aboutness') topic, or the logical and temporal cohesion of events. Crucially, such a discontinuity can involve topical constituents; for instance, when the aboutness topic shifts to another pre-existing topic or when a new narrative section opens. At the same time, discontinuity can also entail unidentifiable Subjects in the case of event-oriented assertions that draw attention to an all-new situation ('thetic' or 'sentence-focal'), and involve clauses that bring new topics into discourse.[68]

3.4 Other variables correlating with VS for identifiable referents

The previous section has identified further distributional tendencies for the apparently 'free' identifiable Subjects, but what about the remaining cases of topical Subjects? Are other variables at play, or does the position of such Subjects remain partly flexible?[69] One other factor is the presence of goals/sources, discussed briefly above in the context of adverbials (Section 2.2.), and a few further remarks follow in this section. However, a wider functional investigation of VS clauses is yet to be undertaken, and should also consider different focus types.[70]

The appearance of some identifiable entities in the post-verbal position can be explained as afterthought. Clear cases of afterthought are those in which the S is separated from the preceding by an intonation pause, indicating the speaker's original intention to end the clause without a lexical argument, as illustrated in (17) below. Elsewhere, however, it is not always easy to identify cases of afterthought; sometimes, the intonation boundary is less clear, or is not indicated in a transcribed source.

---

67 This concept can be traced back to Givón (1983), but is defined somewhat more loosely here. On the more specific concept of referential continuity, see also Section 1.4. above. For further references on 'topic discontinuity', see also Owens et al. (2009: 40–41).

68 As a side note, this notion of theticity usually used for new referents could even be extended to some VS clauses with identifiable characters, thus also creating a meta-category for VS with regard to deixis. For instance, clauses with 'to come'+identifiable character resemble clauses with new referents in that both arguably draw attention to a situation as a whole; in the case of the former, the (re) appearance of a character. This suggestion should be explored further in future studies.

69 Compare this with the claims in Dryer (1997: 88) about VS(/A) as possibly the unmarked or alternative order for Spanish and Polish (both AVO).

70 This will require going beyond the present corpus, as the current data do not include a meaningful number of clauses whose Subject is clearly in narrow focus.

(17) Context: conversation between a ruler and a poor Jew
munìx-le| *ħakom-a*…|
sigh.PRT-3ms ruler-ms
'He sighed, (that is,) the ruler (and said to the poor Jew…)'
(Own fieldwork)

A post-posing effect is also observable with the adverbial *hatxa* 'so, in this way', at least when it is placed clause-initially, as shown in (18a) below. This tendency likely also affects transitive subjects, though only one example thereof is attested in the corpus (in (18b)). In these examples, the post-posing affects identifiable referents, which would otherwise have likely been pre-verbal (certainly in the case of the A). Such XVS(/A) structures serve to bind the modifying adverb to its verb rather than separating them by the topic. They may also be part of the afore-discussed verb-medial syntax (Section 2.2. above), but this should be investigated further.

(18) (a) Context: a man with a wounded leg visited my father, a doctor.
ɪθ-wa-le *parsìmta,*| *hatxa* *pʃil-wā-la* *àql-e.*|
EXIST.PST-POSS.3ms wound thus get.infected.PRT-PST-3fs leg.fs-POSS.3ms
'He had a wound, his leg had become infected like that.'

(b) *zɪl-le* *mɪr-re* *ta* *baxta* *hàtxa* *g-emɪr* *gurg-a.*|
go.PRT-3ms say.PRT-3ms to wife such REAL-say.3ms wolf-ms
'He went and said to his wife, "This is what [the] wolf says".'
(Own fieldwork)

4 The intransitive clause within pragmatic tendencies and word order shifts: a comparative excursus

In the previous discussion, it has been shown that in AVO Neo-Aramaic, the pre-verbal position tends to be restricted to topical Subjects, with a few exceptions. So how does this distribution relate to the general word order profile of Neo-Aramaic, and its pragmatics? For instance, is the distributional divergence between the Agent and the Subject solely an effect of their distinct discourse-pragmatic tendencies, or is transitivity itself also a meaningful variable? Second, are the generalizations based on NENA comparable to the situation in other Neo-Semitic languages which also tend towards AVO? Both questions are briefly considered in this section. I present the data from spoken Neo-Arabic which – like Neo-Aramaic – has been claimed to have pragmatically-driven word order. I include an Arabic variety that is co-territorial

with NENA, and whose speech communities have been in long-standing contact with Aramaic speakers: the Jewish Qəltu dialect of Aqra and Erbil (see Section 1.4. above).

The comparative evidence available indicates that in Neo-Aramaic, both transitivity and pragmatics are more closely linked with word order variation than is the case in Neo-Arabic. Thus, in several varieties of spoken Arabic, VS is statistically more dominant in general, as well as being less functionally restricted than in AVO Neo-Aramaic.

As regards Subject position vis-à-vis the larger syntactic system, I compare different clause types across several distinct Neo-Aramaic dialects. This includes those dialects which have shifted from the original AVO to AOV (see Section 1.1.), thus converging to some degree with the neighbouring Iranian. Such a radical restructuring of the transitive domain could mean that the pre-verbal position in general will lose its topical association in the process. By default, the (immediately) pre-verbal slot now accommodates even unidentifiable, non-topical objects, which often find themselves clause-initially, given that Agents are typically left unexpressed (see Table 3 above). This pragmatic prediction is met by the sample of data. I present findings from five corpora of Neo-Aramaic varieties characterized by increasing degrees of OV rates. In this way, I show that the rise of OV in Neo-Aramaic correlates very closely with the erosion of VS. In fact, Subjects – including unidentifiable ones – show higher rates of pre-verbal placement than other (prototypically) non-topical arguments, which could reflect a diachronic tendency. These preliminary findings identify important typological tendencies and show the relevance of the intransitive clause therefor, which deserve to be investigated further in the future. The present findings should also be compared with corpora of pre-modern (Eastern) Aramaic, where OV also appears in some varieties and text types (see Section 1.1. above). The comparison of the distribution of subjects and objects across different varieties of Eastern Aramaic with attention to these arguments' pragmatics promises to be a worthwhile research avenue.

4.1 Broader VS(/VA): the significance of transitivity and identifiability in Neo-Aramaic in comparison with Neo-Arabic

First, in Neo-Aramaic, is the post-verbal placement of subjects restricted to the intransitive ones, or is this discrepancy between the A and the S simply a side-effect of the Agent's prototypical topicality? The first observation in response to this question is that the post-posing of the A is apparently dispreferred even when this argument *is* new. The three attested unidentifiable Agents (which happen to be accompanied by a lexical object) appear pre-verbally, as illustrated in (19a–b) below. Instead of post-posing, their status as new information is highlighted by their intonational separation, as shown in (19a), which could turn out to be a general tendency for unidentifiable

Agents.[71] In general, therefore, the post-posing of Agents is extremely infrequent in AVO Neo-Aramaic,[72] and is probably due to very specific information-structural functions, such as narrow focus and afterthought (see Section 3.4. above).

(19) AVO Neo-Aramaic

(a) *gòr-a|* *g-ɪb-e* *-wa-la* *raba* *bàxt-e.|*
husband-ms REAL-love-A.3ms -PST.IPFV-O.3fs much wife.fs-POSS.3ms
'A [certain] husband loved his wife very much.'
(Own fieldwork)

(b) Jewish Zakho Neo-Aramaic
raba *raba* *ʿišq* *xil-le* *lɪbb-eni.*
much much desire.ms eat.PRT-3ms heart.PL-POSS.1PL
'A great desire has consumed our hearts.'
(Meehan and Alon 1979: VII: §9)

This situation contrasts with that found in several Arabic varieties. First, Arabic does allow the post-posing of Agents. This applies to J. Qəltu Arabic (Iraqi Kurdistan) as well as other previously-studied varieties. In the tagged corpus of J. Qəltu Arabic, the post-posing rate of Agents stands at 29 % (10/35), the vast majority of which are identifiable. The VA order with a topical Agent is illustrated for J. Qəltu Arabic in (20) below. This distribution reveals a significant difference in the placement of Agents compared to AVO NENA ($p < 0.05$). In J. Qəltu Arabic, the rate of Agent post-posing is even higher than in the previously-studied Jordanian and Peninsular Arabic (Owens et al. 2009; Al-Shawashreh 2016). In fact, in the latter two, the post-posing of Agents is similarly rare to that in AVO NENA, which further emphasises the distinctiveness of AVO NENA from its co-territorial Qəltu Arabic.[73]

(20) Qəltu Arabic, Jewish Aqra and Erbil
az *fad garre,* *ʿallm-ūw-a* *ban-āt* *ɪl-mɪsɪlm-īn…*
so once teach.PRT-A.3PL-O.3fs girl-PL DEF-Muslim-PL
'So once, the Muslims girls taught her (to say…)'
(Jastrow 1990: 132: §160)

---

71 This cannot be verified for (15b), which comes from a text. Compare this strategy with detachment ('left dislocation') constructions, in which a newly-activated topic is placed just before the clause, and usually resumed from the clause proper by a pronoun (see Lambrecht 1994: 181–182).

72 For the discussion on 'agent-first' (with 'agent' understood semantically) with a focus on western Austronesian, see Riesberg et al. (2019), and the references therein.

73 The rates of VA reported for Peninsular and Jordanian are, respectively, 12 % (Owens et al. 2009: 48) and 14 % (Al-Shawashreh 2016: 98). I calculated the latter rate by adding the rates of clauses with and without objects, reported separately (ibid). The (otherwise very in-depth) study of Neo-Arabic by Dahlgren (1998) unfortunately does not consider transitivity as a variable.

Table 9: The distribution of identifiable subjects in J. Qəltu Arabic versus AVO NENA.

	SV	VS	Total *n*
J. Qəltu Arabic	51 % (18)	49 % (17)	35
AVO NENA	55 % (53)	45 % (44)	97
Total *n*	71	61	132

$p > 0.05$.

Since the majority of Agents are identifiable to begin with, this finding suggests that (Qəltu) Arabic word order is less sensitive to pragmatics, at least as regards the post-verbal position. In J. Qəltu Arabic, this position is not restricted to new material, neither is it more characteristic of 'discourse-discontinuous' topics.

This hypothesis is supported by the situation in the intransitive domain in J. Qəltu Arabic. In the corpus, identifiable referents show similar post-verbal rates as in AVO NENA (Table 9).

However, unlike in NENA, the post-verbal position in J. Qəltu Arabic seems to be common with various types of Subjects and verbal lexemes. It thus features many topics which do not mark discourse switches in the sense described in Section 3.3. This is shown in (21), which describes a narrative foreground event:

(21) Qəltu Arabic, Jewish ʿAqra and Arbēl
habl-at *mar-t* *abūy-i…*
conceive.PRT-3fs wife-fs.GEN father.ms-POSS.1SG
'(She – out of fear,) the wife of my father got pregnant.'
(Jastrow 1990: 174: §308)

Overall, therefore, the situation in J. Qəltu Arabic points to a greater flexibility and higher frequency of the 'verb-subject' permutation, relatively to NENA. It occurs relatively more commonly with all definite subjects, including all types of topical S and A arguments. At the same time, what still seems to hold across NENA and Arabic is that the pre-verbal position (clause-initial or otherwise) is strongly associated with topical material. However, a larger and more closely-annotated corpus would be required to corroborate this.[74]

74 Based on the data available (mine and published elsewhere), comparison remains difficult. For instance, in a sub-set of the Peninsular corpus (Owens et al. 2009), roughly 30 % of 'new' nouns are pre-verbal (Owens et al., 64, footnote 20), but this includes both transitive and intransitive subject.

4.2 Erosion of VS and the rise of AOV: comparison with AOV Neo-Aramaic

The above-discussed 'pre-verbal–topical' association fits the general AVO profile of Jewish Duhok–Zakho Neo-Aramaic, and, more specifically, with the 'topic-comment' information structure underpinning its word order. This information-structural generalization about 'topic before the predicate' is thus applicable to other clause types. As shown in Table 10 below (see also 3.1. above), this tendency is especially strongly manifested in existential clauses and on the possessees of predicative possession. Both arguments are (almost) always unidentifiable, and both clause types are used to bring new referents into the discourse. Such clauses have the highest rates of post-verbal arguments relatively to other clause types, across most NENA dialects (see Table 10).

This 'topic before the predicate' tendency also manifests itself in objects in AVO NENA, where identifiable (definite) objects appear pre-verbally more often than their indefinite counterparts (Noorlander and Molin 2022: 242; Noorlander et al. forthcoming). This discrepancy between identifiable and unidentifiable objects applies (almost) always in NENA and its co-territorial Arabic, regardless of the overall rates of OV in a given variety (Noorlander et al. forthcoming). Such a pre-verbal shifting of identifiable objects has been interpreted as an increase in topicalization. At the same time, it eventually leads to the rise of OV as the dominant and thus unmarked order (Noorlander et al. forthcoming). This, in turn, apparently results in the erosion of the topicalizing function of the (immediately), as (A)OV becomes entrenched for all object types.

So does this loss of topicality restriction on the (immediately) preverbal constituent in NENA also apply to Subjects? This section compares the statistically dominant word orders across several clause types, and across four NENA varieties of distinct transitive profiles (see Sections 1.1.&1.4.). The data in Table 10 demonstrate that the rise of OV leads to the loss of this topical restriction on the pre-verbal position, which does manifest itself across clauses. At the same time, different arguments shift to the pre-verbal position at varying rates, as discussed further below. This disparity is clearly visible in Figure 2, which compares the rates of post-verbal objects and of intransitive subjects specifically.

The Jewish variety of Sena is the simplest to analyse, because it reflects the final stage of the shift to OV. Here, the 'shifting forward' of objects correlates with a near-obligatory pre-verbal placement for all arguments, including all types of Subjects, existees and possessees, apparently due to the total bleaching of the (immediately) pre-verbal slot of its topical association. This distribution makes AOV Aramaic syntactically much more rigid than the dialects tending towards AVO (cf. Noorlander and Molin 2022: 244–246). Unlike the latter, the former cannot use word order for pragmatic nuance, or in a free manner. Interestingly, the pragmatic uniformity exhibited by Jewish Sena NENA (which enforces the structure '(topic-)non-topic-

Table 10: Post-predicate rates of nominal arguments across clause types. The presence of other arguments is not considered here. Sources: The count data for objects and possessees in Jewish Sena and Betanure and Christian Urmia come from Noorlander et al. (forthcoming), the other counts are the work of the present author.

General tendency (sorted from AVO to AOV)	Identifiable subjects	Subjects (all)	New subjects	Identifiable objects[a]	Objects (all)	New objects	Existential ('There is_X.')	Possessee in predicative possession
AVO (J. Duhok and Zakho)	**41 %**	**51 %**	**74 %**	**73 %**	**87 %**	**96 %**	**96 %**	**94 %**
	46/112	83/164	37/50	45/62	182/210	161/168	49/51	34/36
Tending to AVO (J. Betanure)	**12 %**	**26 %**	**39 %**	**53 %**	**76 %**	**92 %**	–	–
	6/50	26/101	20/51	48/90	**159/210**	110/116		
Tending to AOV (C. Urmia)	**10 %**	**15 %**	**21 %**	**8 %**	**17 %**	**40 %**	**68 %**	**79 %**
	5/48	12/81	7/33	12/153	24/144C	17/43	31/46	19/24
AOV (J. Sena)	**5 %**	–[b]	–	**12 %**	**5 %**	**3 %**	**0 %**	**2 %**
		8/158		10/86	18/331	7/244	0/80	1/65

[a]Objects of monotransitive ('Patients') and Themes of ditransitive. [b]I have not checked the specific rates of post-verbal identifiable and unidentifiable Subjects separately, since the total frequency of all Subjects is already extremely low. The percentages are in bold.

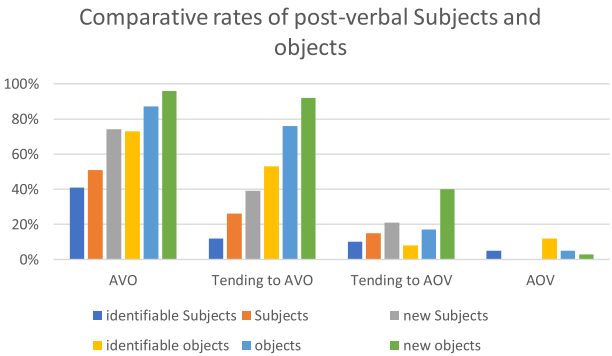

Figure 2: Post-verbal rates of subjects and objects from a different types of Neo-Aramaic.

predicate' across clause types) is only one possible (re)configuration of word order. Latin, for instance, makes productive use of the VS construction (cf. Bolkestein 1995), despite its general AOV profile. In other words, while topical and non-topical objects apparently tend to be pre-verbal, subjects show more variability (Devine and Stephens 2006: e.g. 150–51). According to Bolkestein (1995), moreover, Latin apparently has an even greater tolerance for topical Subjects to be post-verbal than does Italian, though the latter does place objects post-verbally. In NENA, by contrast, the shift of AVO > AOV apparently correlates with a *diminishing* functionality of the post-verbal slot.[75] A closer comparative and diachronic study of different AOV systems is required to understand their differences properly.

Returning to the dialects in Table 10, in the varieties whose shift to OV is less complete, the distribution of non-topical arguments is more complex. First, pre-verbal placement is more likely with *identifiable* Subjects in both Jewish Betanure and the Christian Urmia dialect, as also reported for objects (cf. Noorlander et al. forthcoming). In the former dialect, for instance, the post-verbal rates for unidentifiable Subjects are 39 %, but only 12 % for topical Subjects. At the same time, the significance of identifiability as a variable is lower in the more strongly AOV variety, which could reflect a diachronic tendency. Thus, identifiability is still somewhat significant in Jewish Betanure ($p = 0.003$), but not for Christian Urmia ($p > 0.05$).

Despite the relatively lower rates of pre-verbal placement of Subjects relative to objects, *unidentifiable* subjects are already very commonly pre-verbal in the 'in between' NENA varieties. This applies even to Jewish Betanure (47 % pre-verbal), which still tends to AVO. This rate represents a significant drop in the post-verbal placement of Subjects relatively to the dialects of Duhok and Zakho (27 % pre-verbal), even though all three belong to the same dialect cluster (e.g. Molin 2024; Mutzafi 2008a). As

75 But see Lambrecht (2000: 648–49) for Italian word order more generally.

shown above, in Duhok/Zakho, unidentifiable Subjects are by default post-verbal, especially in sentence-focal/thetic constructions (83 %), except for specific syntactic conditions (see Section 2.2. above). Examples of sentence-focal constructions for Jewish Betanure, Jewish Duhok/Zakho and Christian Urmia are given in (22a), (22b) and (22c), respectively. The near-minimal pair with regard to word order shows that even in sentence-focal constructions, the SV order has already become more dominant than VS in Jewish Betanure:

(22) Jewish Betanure NENA, partly AVO
(a) *bət səθwa talg-a k-awəš-wa.*
in winter snow-ms REAL-descend.3ms-PST
'In winter, SNOW would fall.'
(Mutzafi 2008a: 144, §38)
(b) Jewish Duhok NENA, AVO
bət səθwa k-aw-e-wa tàlg-a.|
in winter REAL-be-3ms-PST snow-ms
'In winter, there would be SNOW.'
(Own fieldwork)
(c) Christian Urmia NENA, partly AOV
àxun-i,| *nùyr-a pílɪ -va…*|
brother.ms-POSS.1SG fire-ms PTCP.fall.ms -COP.PST.3ms
'(He said) "my brother, FIRE broke out…"'
(Khan 2016_{IV}: 198)

This difference between Betanure and Duhok/Zakho in terms of Subjects correlates with their difference in objects. More specifically, in the former, roughly half of the definite objects are now pre-verbal, but post-verbal ones are still strongly post-verbal. This means that the pre-verbal placement of Subjects – including indefinite ones – is more common than that of objects, including indefinites (more on this below). In terms of grammaticalization, therefore, it appears that even this incipient forward movement of objects (mostly affecting definites) works to attract non-topical arguments to the pre-verbal slots.

In the NENA variety tending towards AOV (Christian Urmia), unidentifiable referents are also more strongly pre-verbal, and in this case overwhelmingly so. This was shown in Table 10 above, and illustrated in (22c), which has a thetic/sentence-focal construction.[76] As regards the reverse, i.e. post-verbal position in Christian Urmia, it is used with both human and inanimate referents, and with both

76 Still, the 'all-new information' status in such SV clauses is formally flagged by nucleus stress on the Subject. This makes clauses like in (22c) both functionally and formally similar to subject-accented (i.e. thetic) constructions in English or German (cf. Sasse 1987: 520–526).

identifiables and unidentifiables. This cautions us against proposing a single discrete 'functional stronghold' for the last stage of the VS order. In Christian Urmia, this stronghold accommodates different types of 'survivors'.

As regards the different rates of shifting across clauses and dialects more broadly, I already observed that indefinite objects show somewhat slower rates of pre-verbal shifting than non-topical subjects. This difference is statistically significant for Neo-Aramaic varieties that still tend to AVO (Jewish Duhok/Zakho and Betanure), but not for the corpora tending towards AOV. This situation suggests that in addition to information structure, argument structure also plays a role in the diachronic change. Every clause is to some extent governed by its own syntactic tendencies, which likely involve the prototypical function of the structure in question, and its typical number of constituents.[77]

This view is further supported by the behaviour of the two non-verbal clauses shown in Table 10 above, particularly in the Christian NENA of Urmia. Despite this variety's strong tendency towards AOV (60 % pre-verbal among new objects), Christian Urmia still makes productive use of the post-verbal slot for two non-topical arguments: existees (68 % post-predicate) and possessees (79 % post-predicate). The dominant order in the existential clause is given in (23a), while (23b) illustrates the innovative 'existee-predicate' order. Both examples come from the opening of a story.

(23) Christian Urmia Neo-Aramaic

(a) *ìt-va*| *ʿdān-ḵ̂ădimì*| *mara* *ɪt-va* *l-ì-tva*| *ɪt-va* *xa-dana*
EXIST-PST time-ancient say.INF EXIST-PST NEG-EXIST-PST EXIST-PST INDEF-time
màlc-a| *Sanxìro.*|
king-ms Sankhiro
'In times of old, there was, they say, there was there was not, there was a king, Sankhiro (the king of the Assyrians).'
(Khan 2016$_{IV}$: 2: §1)

(b) *xà-dana*| *catʃ͡ala*| *ìt-va*| *ɟu-xà*| *mdìta.*|
INDEF-time bald EXIST-PST in-INDEF city
'Once, there was a bald man in a town.'
(Khan 2016$_{IV}$: 26: §1)

5 Conclusions

This paper is a study of word order tendencies across several varieties of Northeastern Neo-Aramaic – a group of contemporary vernacular varieties of Aramaic,

77 Ultimately, however, all clauses across the NENA corpora are affected by the changes taking place in the transitive domain.

which is part of the Northwest Semitic subgroup. The present study thus contributes to the field of corpus-based typology and syntax within Neo-Semitic, which is still only incipient. More broadly, this investigation also demonstrates the significance of corpus data for understanding the usage-based effects of semantic and pragmatic differences between different arguments, such as the intransitive versus transitive subjects. At the same time, I also show that different arguments (such as intransitive subjects, objects and transitive subjects) are capable of asserting some behavioural independence, even while sharing their semantic or pragmatic features with other argument types. Overall, therefore, semantics, information and argument structure alike are important for understanding the variation in (Aramaic) word order.

At the heart of this study are two closely related dialects of Northeastern Neo-Aramaic characterized by conservative syntax, which is manifested in their maintenance of an erstwhile AVO tendency. They are spoken by the Jews of Duhok and Zakho in Northwestern Iraq and are now critically endangered. The distribution of the intransitive versus transitive subjects in this dialect challenges the notion of a single, unified 'subject' category in Neo-Aramaic, as these two subjects have radically different distributions. The Agent is overwhelmingly pre-verbal, even on the rare occasion that it is unidentifiable, while the S splits roughly 'down the middle' between SV and VS.

Moreover, it is identifiable subjects in particular which display a great degree of positional freedom. Unidentifiable Subjects prefer the post-verbal position – a tendency which is especially strong if no goal is present in the clause (83 % of VS in such clauses). By contrast, identifiable referents are roughly equally common in both positions under the same circumstances (45 % in VS). This, in turn, calls for a multivariable account of word order variation, while allowing for the possibility of free distribution.

The distribution of Subjects in Neo-Aramaic indicates that pre-verbal positions are generally reserved for identifiable constituents, which fits with the general TOPIC-COMMENT structure of AVO Neo-Aramaic. By contrast, the following clauses are overwhelmingly VS: those with all-new status ('sentence-focus'), those which draw attention to a (new) situation ('thetic'), and which introduce new referents into the discourse. Moreover, the distribution of identifiable referents is not entirely free, since VS coincides with particular discourse functions. Many of such clauses indicate some sort of 'discourse discontinuity': they reintroduce or switch a topic, connect two discourse sections or begin a new one. This interpretation is supported by a number of lexical, syntactic and discourse-based correlates, but cannot be quantified through referential distance. I proposed to subsume most of these clauses under 'discourse discontinuity', broadly understood here as referring to 'seams' in the narrative flow. Moreover, the usefulness of such a meta-category lies in its ability to unite many functions of VS, subsuming identifiable and unidentifiable referents alike: topic

introductions, reintroductions and switches, or eventive assertions. Thus, Neo-Aramaic Subject variation correlates especially closely with information structure and discourse pragmatics. On the other hand, this variation cannot be reduced to simple semantic alignment, as the post-verbal, object-like position is available to agentive and non-agentive ('unaccusative') Subjects alike.

The comparative excursus (Section 4.1) further sharpens our understanding of Neo-Aramaic word order, by comparing it with its co-territorial (Jewish Qəltu) Arabic. While Neo-Aramaic strongly disprefers the post-posing of Agents, in Jewish Qəltu Arabic, post-verbal *transitive* subjects are more frequent. Moreover, post-verbal intransitive subjects do not show the same kind of sensitivity to information structure and discourse found for Neo-Aramaic.

In the comparison of clause types across four different Neo-Aramaic varieties offered in Section 4.2., the doculects are aligned on a spectrum from AVO to AOV. This demonstrated the relevance of the intransitive clause for understanding larger-scale variation in word order and information structure. In particular, in Neo-Aramaic, the increased rates of OV drastically reduce the productivity of the VS clause. This productivity drops significantly even before the shift to OV is complete. For instance, in a dialect where 76 % of objects are still post-verbal, most Subjects are already pre-verbal (only 26 % post-verbal), including the majority of unidentifiable ones. This suggests that even in an incipient shift to OV, the (immediately) pre-verbal position has already lost its exclusive topical association, since it now allows many non-topical subjects. This seems to suggest that a change in frequency in Neo-Aramaic is accompanied by a change in the information structure underlying syntax, which highlights the significance of both information-structural and statistical approaches to word order. These findings should be further explored in the future and compared with the situation in other languages, and indeed with historical varieties of Aramaic. Finally, the cross-dialectal Neo-Aramaic data also compare the pre- and post-verbal rates for possessees in predicative possession, and for existees. These arguments seem to 'hold on' to the post-predicate slot especially 'strongly', in comparison to Subjects and Objects. Each clause must thus also be considered a system in its own right.

Special symbols and abbreviations

|	intonation boundary
◌̀	Nucleus stress (main/major stress within an intonation unit)
C.	Christian (doculect)
J.	Jewish (doculect)

References

Abbott, Barbara. 1997. Definiteness and existentials. *Language* 73(1). 103–108.

Al-Shawashreh, Ekab. 2016. *Aspects of grammatical variation in Jordanian Arabic*. Doctoral dissertation. Université d'Ottawa/University of Ottawa.

Andersen, Torben. 1988. Ergativity in Pari, a nilotic OVS language. *Lingua* 75(4). 289–324.

Andrews, Avery. 2007. The major functions of the NP. In T. Shopen (ed.), *Language typology and syntactic description*, 2nd edn. 1, 62–154. Cambridge: Cambridge University Press.

Arnaiz, Alfredo. 1997. An overview of the main word order characteristics of Romance. In Anna Siewierska (ed.), *1997b, constituent Order in the Languages of Europe*, 47–74. Berlin: Mouton De Gruyter.

Bickel, Balthasar. 2003. Referential density in discourse and syntactic typology. *Language* 708–736. https://doi.org/10.1353/lan.2003.0205.

Bickel, Balthasar. 2008. On the scope of the referential hierarchy in the typology of grammatical relations. In Greville Corbett & Michael Noonan (eds.), *Case and grammatical relations: Studies in honor of Bernard Comrie*, 191–210. Amsterdam: John Benjamins.

Bickel, Balthasar. 2011. Statistical modeling of language universals. *Linguistic Typology* 15(2). 401–413.

Bickel, Balthasar, Alena Witzlack-Makarevich & Taras Zakharko. 2015. Typological evidence against universal effects of referential scales on case alignment. In AndrejIna Bornkessel-Schlesewsky, Andrej Andrej Malchukov & &Marc D. Richards (eds.), *Scales and hierarchies: A cross-disciplinary perspective*, 7–44. Berlin: Mouton De Gruyter.

Bolkestein, Machtelt. 1995. Functions of verb-subject order in Latin. *STUF – Language typology and universals* 48(1–2). 32–43.

Chafe, Wallace. 1987. Cognitive constraints on information flow. In Russell S. Tomlin (ed.), *Coherence and grounding in discourse*, 11, 2–51. Amsterdam: John Benjamins.

Coghill, Eleanor. 2014. Differential object-marking in Neo-Aramaic. *Linguistics* 52. 335–364.

Coghill, Eleanor. 2016. *The rise and fall of ergativity in Aramaic: Cycles of alignment change*. Oxford: Oxford University Press.

Coghill, Eleanor. 2018. Information structure in the Neo-Aramaic dialect of Telkepe. In Evangelia Adamou, Katharina Haude & Martine Vanhove (eds.), *Information structure in lesser-described languages: Studies in prosody and syntax*, 298–327. Amsterdam: John Benjamins.

Cohen, Eran. 2012. *The syntax of Neo-Aramaic: The Jewish Dialect of Zakho. Gorgias Neo-Aramaic studies*. Piscataway: Gorgias.

Comrie, Bernard. 1978. Ergativity. In W. Lehmann (ed.), *Syntactic typology: Studies in the phenomenology of language*, 329–393. Sussex: The Harvester Press.

Comrie, Bernard. 1988. Topics, grammaticalized topics, and subjects. *Berkeley Linguistics Society* 14. 265–279.

Cook, Edward M. 1986. Word order in the Aramaic of Daniel. *Afroasiatic Linguistics* 9. 111–126.

Croft, William. 2012. *Verbs: Aspect and causal structure*. Oxford: Oxford University Press.

Dahlgren, Sven-Olof. 1998. *Word order in Arabic*. Seychelles: Acta Universitatis Gothoburgensis.

Devine, Andrew M. & Laurence D. Stephens. 2006. *Latin word order: Structured Meaning and information*. Oxford/New York: Oxford University Press.

Dixon, Robert. 1979. Ergativity. *Language* 55. 59–138.

Dixon, Robert M. W. 1994. *Ergativity*. Cambridge: Cambridge University Press.

Donohue, Michael. 2008. *Semantic alignment systems: What's what, and what's not?* In Mark Donohue & Søren Winchmann (eds.), *The typology of semantic alignment*, 24–75. Oxford: Oxford University Press.

Dryer, Matthew. 1997. On the 6-way word order typology. *Studies in Language* 21. 69–103.
Dryer, Matthew. 2007. Word order. In Timothy Shopen (ed.), *Language Typology and syntactic description*, 61–131. Cambridge: Cambridge University Press.
Du Bois, John. 1987a. Absolutive zero. *Lingua* 71(2). 203–222.
Du Bois, John D. 1987b. The discourse basis of ergativity. *Language* 63(4). 805–855.
Folmer, Margaretha. 1995. *The Aramaic language in the Achaemenid period: A study in linguistic variation*. Leuven: Peeters.
Garbell, Irena. 1965. *The Jewish Neo-Aramaic Dialect of Persian Azerbaijan*. The Hague: Mouton.
Givón, Talmy. 1983. Topic continuity in discourse: An introduction. In Talmy Givón (ed.), *Topic Continuity in discourse: A quantitative cross-language study*, 1–41. Amsterdam: John Benjamins.
Gutman, Ariel. 2018. *Attributive constructions in North-Eastern Neo-Aramaic*. Berlin: Language Science Press.
Gzella, Holger. 2015. *A cultural history of Aramaic: From the beginnings to the advent of Islam*. Leiden: Brill.
Haig, Geoffrey. 2008. *Alignment change in Iranian languages*. Berlin: De Gruyter.
Haig, Geoffrey & Geoffrey Khan (eds.). 2019. *The languages and linguistics of Western Asia*. Berlin: Mouton De Gruyter.
Haig, Geoffrey & Stefan Schnell. 2016. The discourse basis of ergativity revisited. *Language* 92(3). 1–14.
Haig, Geoffrey, Stefan Schnell & Nils Schiborr. 2022. Universals of reference in discourse and grammar: Evidence from the Multi-CAST collection of spoken corpora. In Geoffrey Haig, Stefan Schnell & Frank Seifart (eds.), *Doing corpus-based typology with spoken language corpora: State of the art*, 141–177. Honolulu: University of Hawai'I Press.
Haig, Geoffrey, Mohammad Rasekh-Mahand, Donald Stilo, Laurentia Schreiber & Nils Schiborr (eds.), forthcoming. *Post-predicate elements in the Western Asian transition zone: A corpus-based approach to areal typology*. Berlin: Language Science Press.
Harris, Alice C. & Lyle Campbell. 1995. *Historical syntax in cross-linguistic perspective*. Cambridge: Cambridge University Press.
Haspelmath, Martin. 2011. On S, A, P, T, and R as comparative concepts for alignment typology. *Linguistic Typology* 15(3). 535–567.
Hewitt, Steve. 2006. Arabic: Verb-subject-object or verb-given-new? Implications for word order typology. In *Conference on Communication and information Structure in spoken Arabic*.
Hoberman, Robert D. 1989. *The syntax and semantics of verb morphology in modern Aramaic : A Jewish dialect of Iraqi Kurdistan*. New Haven: American Oriental Society.
Holes, Clive. 1995. *Modern Arabic: Structures, functions, and varieties*. London: Longman.
Jacennik, Barbara & Matthew Dryer. 1992. Verb-subject order in Polish. In Doris Payne (ed.), *Pragmatics of word order flexibility*, 209–241. Amsterdam: John Benjamins.
Jastrow, Otto. 1990. *Der Arabische Dialekt der Juden von ʿAqra and Arbīl*. Wiesbaden: Harrasowitz.
Kaufman, Stephen A. 1974. *The Akkadian influences on Aramaic*. Chicago and London: The University of Chicago Press.
Keenan, Edward L. 1976. Towards a universal definition of subject. In Charles Li (ed.), *Subject and topic*, 303–333. Cambridge, MA: Academic Press.
Khan, Geoffrey. 2007. The Northeastern Neo-Aramaic dialects. *Journal of Semitic Studies* 52(1). 1–20.
Khan, Geoffrey. 2008. *The Neo-Aramaic dialect of Barwar. Volume I: Grammar*. Leiden: Brill.
Khan, Geoffrey. 2009. *The Jewish Neo-Aramaic Dialect of Sanandaj*. Piscataway: Gorgias Press.
Khan, Geoffrey. 2016. *The Neo-Aramaic dialect of the Assyrian Christians of Urmi. Volume I: Grammar: Phonology and Morphology, Volume II: Syntax, Volume IV: Texts*. Leiden: Brill.
Khan, Geoffrey & Masoud Mohammadirad. 2024. *Language contact in Sanandaj: A study of the impact of Iranian on Neo-Aramaic*. Berlin: De Gruyter.

Kutscher, Yehezqel. 1952. *'Aramit miqra'it – 'aramit mizrahit hi' o ma'aravit?*[Biblical Aramaic – Eastern Aramaic or Western Aramaic?] *World Congress of Jewish studies, Summer 1 1947,1*, Vol. 1, 123–127. Jerusalem.

Lambrecht, Knud. 1994. *Information structure and sentence form: Topic, focus, and the mental representations of discourse referents*. Cambridge: Cambridge University Press.

Lambrecht, Knud. 2000. When subjects behave like objects. *Studies in Language* 24(3). 611–682.

Lambrecht, Knud & Maria Polinsky. 1997. Typological variation in sentence-focus constructions. *CLS* 33(2). 141–165.

Levin, Beth & Malka Rappaport Hovav. 1992. The lexical semantics of verbs of motion: The perspective from unaccusativity. In Iggy Roca (ed.), *Thematic structure: Its role in grammar*. Berlin: Mouton De Gruyter.

Levshina, Natalia. 2015. *How to do linguistics with R: Data exploration and statistical analysis*. Amsterdam: John Benjamins.

Matras, Yaron & Hans-Jürgen Sasse. 1995. VS-order and theticity in European languages. *STUF - Language Typology and Universals* 48(1–2). 1–31.

Meehan, Charles & Jacqueline Alon. 1979. The Boy whose Tunic Stuck to him: A Folktale in the Jewish Neo-Aramaic dialect of Zakho (Iraqi Kurdistan). *Israel Oriental Studies* IX. 174–202.

Molin, Dorota. 2024. *The Jewish Neo-Aramaic dialect of Dohok: A comparative-typological study*. Leiden: Brill.

Mutzafi, Hezy. 2008a. *The Jewish Neo-Aramaic dialect of Betanure (Province of Dihok)* Wiesbaden: Harrasowitz.

Mutzafi, Hezy. 2008b. Trans-zab Jewish Neo-Aramaic. *BSOAS Bulletin of the School of Oriental and African Studies* 1(3). 409–431.

Naro, Anthony J. & Sebastião J. Votre. 1999. Discourse motivations for linguistic regularities: Verb/subject order in spoken Brazilian Portuguese. *Probus* 11. 75–100.

Noorlander, Paul M. 2014. Diversity in convergence: Kurdish and Aramaic variation Entangled. *Journal of Kurdish Studies* 2. 201–224.

Noorlander, Paul M. 2021. *Ergativity and other alignment types in Neo-Aramaic*, vol. 79, 708–736. Leiden: Brill.

Noorlander, Paul M. & Dorota Molin. 2022. Word order typology in North Eastern Neo-Aramaic: Towards a corpus-based approach. In Hiwa Asadpour & Thomas Jügel (eds.), *Word order variation: Semitic, Turkic and Indo-European languages in contact (Studia Typologica [STTYP])*, 235–258. Berlin: De Gruyter.

Noorlander, Paul M., Dorota Molin & Geoffrey Haig. Forthcoming. Between VO and OV in Arabic and Aramaic: A corpus-based typology with implications for word-order shifts. *Studies in Language*, In press.

Noorlander, Paul M. Forthcoming. Neo-Aramaic in Iran and northeastern Iraq. In *Post-predicate elements in the Western Asian transition zone: A corpus-based approach to areal typology*.

Owens, Jonathan & Robin Dodsworth. 2009. Stability in subject-verb word order: From contemporary Arabian Peninsular Arabic to Biblical Aramaic. *Anthropological Linguistics* 51(2). 151–175.

Owens, Jonathan, Robin Dodsworth & Trent Rockwood. 2009. Subject-verb order in spoken Arabic: Morpholexical and event-based factors. *Language Variation and Change* 21(1). 39–67.

Prince, Ellen F. 1981. Towards a taxonomy of given/new information. In Peter. Cole (ed.), *Radical pragmatics*. New York: Academic Press.

Riesberg, Sonja, Kurt Malcher & Nikolaus P. Himmelmann. 2019. How universal is agent-first? Evidence from symmetrical voice languages. *Language* 95(3). 523–561.

Sasse, Hans-Jürgen. 1987. The thetic categorical distinction revisited. *Linguistics* 25. 511–580.

Schnell, Stefan, Nils Schiborr & Geoffrey Haig. 2021. Efficiency in discourse processing: Does morphosyntax adapt to accommodate new referents? *Linguistics Vanguard* 7(s3). 20190064.
Siewierska, Anna. 1988. *Word order rules*. London: Croom Helm.
Siewierska, Anna. 1997b. Variation in major constituent order; a global and a European perspective in *Constituent order*. In Anna Siewierska (ed.), *Constituent order in the languages of Europe*, 475–552.
Siewierska, Anna & Dirk Bakker. 2009. Case and alternative strategies: Word order and agreement marking. In Andrej Malchukov & Andrew Spencer (eds.), *The Oxford handbook of case*, 290–303. Oxford: Oxford University Press.
Tortora, Christina M. 1999. The post-verbal subject position of Italian unaccusative verbs of inherently directed motion. *Amsterdam studies in the theory and history of linguistic science series* 4. 283–298.
Ueno, Mieko & Maria Polinsky. 2009. Does headedness affect processing? A new look at the VO–OV contrast. *Journal of Linguistics* 45(3). 675–710.