

Pregroup Analysis of Persian Sentences

Mehrnoosh Sadrzadeh

School of Electronics and Computer Science

University of Southampton

ms6@ecs.soton.ac.uk

1 Introduction

Pregroups are mathematical structures introduced by Lambek in [17] as replacement for his *type categorial grammars* [18], much used in Linguistics e.g. see [22]. Pregroups have been used to analyze the sentence structure of many languages, for example English [15], French [3], Arabic [2], German [16], Italian [7], Polish [19], and Japanese [6]. The mathematical and logical properties of pregroups have been studied in [4, 5, 14, 23]. In this paper, we present a pregroup analysis of Persian grammar¹.

We start with a brief introduction to pregroups and then analyze the Persian simple and compound sentences with simple tense verbs and explicit subjects and objects. Our references to Persian grammar are [8, 21] in English and [20, 10] in Persian. In the course of analysis, we use a *reduction diagram* based on the diagrammatic toolkit of compact closed categories² originally developed in [12] and used for quantum computation in [1]. Based on these diagrams we introduce a *degree of nesting*, which assigns to each sentence a number that reflects the complexity of its analysis. These degrees can be used to compare the complexity of different sentences in a language and the same sentence in different languages. As further examples, we use our results to analyze one verse of the Rubaiyat of Omar Khayyam [13, 9] and also two verses of a poem by Hafiz [11]. Extending our analysis to sentences with implicit subject and object pronouns and verbs in compound tenses constitutes future work.

2 Theoretical Background

A pregroup P is a partially ordered non-commutative monoid $(P, \cdot, 1, \leq, (-)^l, (-)^r)$ where each element $p \in P$ has both a left adjoint p^l and a right adjoint p^r . A partially ordered monoid is a set P with a reflexive, transitive, anti-symmetric relation $\leq \subseteq P \times P$ and a binary operation $\cdot: P \times P \rightarrow P$ that preserves the partial order, that is

$$p_1 \leq p_2, \quad p_3 \leq p_4 \implies p_1 \cdot p_3 \leq p_2 \cdot p_4$$

¹The author understands that a combinatorial analysis of Persian sentences with a free word order has been presented in [24]. The free word order is a point of view not shared by our references of Persian grammar [8, 21, 20, 10].

²Indeed, a pregroups is a thin non-symmetric compact closed category. For a bicategorical approach to pregroups see [23].

The non-commutativity of the monoid multiplication means that the order of the operation is important, so in general the following does not hold

$$p_1 \cdot p_2 = p_2 \cdot p_1$$

The multiplication has a unit 1, that is $p \cdot 1 = 1 \cdot p = p$. Each element having a left and a right adjoint means that for each element $p \in P$ there are two other elements $p^l, p^r \in P$ for which we have the following four inequalities

$$\begin{array}{ll} p^l \cdot p \leq 1 & p \cdot p^r \leq 1 \\ 1 \leq p \cdot p^l & 1 \leq p^r \cdot p \end{array}$$

The first two inequalities are referred to as *contractions* and the other two as *expansions*. One can show that the adjoints are unique and that the unit is self adjoint (for proof see for example [14]), that is

$$1^l = 1 = 1^r$$

For the linguistic analysis, one fixes a set of basic types, for example π for subject, o for object, s for statement, etc, and generates a free pregroup over these basic types. The monoid multiplication is the juxtaposition of these types and the unit of multiplication is the empty type that has no effect in juxtaposition. The pregroup will have basic types π, o, s , other simple types such as $\pi^l, \pi^r, o^l, o^r, s^l, s^r$, and compound types such as $(\pi^r \cdot s \cdot o^l)$. For simplicity, we drop the multiplication sign and for example write $(\pi^r s o^l)$ for $(\pi^r \cdot s \cdot o^l)$. The juxtaposition of types of the parts of a grammatically correct sentence has to be less than or equal to its desired type, e.g. statement or question. For example, the types of the parts of a simple English sentence with a transitive verb are as follows

$$\begin{array}{lll} \text{subject} & \text{verb} & \text{object} \\ \pi & (\pi^r s o^l) & o \end{array}$$

The following inequalities hold for the juxtaposition of the above types

$$\pi \pi^r s o^l o \leq 1 s o^l o = s o^l o \leq s 1 = s$$

This means that our sentence reduces to the type s . This is denoted as follows

$$\begin{array}{lllll} \text{subject} & \text{verb} & \text{object} & \rightarrow & \text{sentence} \\ \pi & (\pi^r s o^l) & o & \rightarrow & s \end{array}$$

where \rightarrow stands for the partial order \leq . We call the above procedure, *reduction* or *typing* of the sentence. It is shown in [17, 5] that contraction alone suffices for analysis of sentences. This result gives rise to a decision procedure for reduction of sentences in the pregroup analysis of a language. Some other properties of pregroups that are useful in the linguistic analysis are as follows

1- The adjoint of multiplication is the multiplication of adjoints but in the reverse order

$$(p \cdot q)^l = q^l \cdot p^l \quad (p \cdot q)^r = q^r \cdot p^r$$

2- The adjoint operation is order reversing

$$\frac{p \rightarrow q}{q^r \rightarrow p^r} \quad \frac{p \rightarrow q}{q^l \rightarrow p^l}$$

3- Composition of the opposite adjoints is identity

$$(p^l)^r = (p^r)^l = p$$

4- The adjoint operation is not involutive

$$p^{ll} = (p^l)^l \neq p, \quad p^{rr} = (p^r)^r \neq p$$

This leads to existence of iterated adjoints [14], so each element of a pregroup can have countably many iterated adjoints

$$\dots, p^{ll}, p^l, p, p^r, p^{rr}, \dots$$

3 Typing the Sentence

We start our analysis by the following basic types

1. π_k to k-th person pronoun ($k = 1, 2, 3$ for singular and $k = 4, 5, 6$ for plural)
2. s_j to declarative sentence in j-th simple tense: s_1 for present, s_2 for past, and s_3 for subjunctive
3. σ_j to sentential complement in j-th simple tense
4. q_j to questions in the j-th simple tense
5. t_l to l 'th tense stem, so t_1 to the present stem and t_2 to the past stem
6. o to direct object
7. w to prepositional phrase
8. n to singular noun
9. p to plural noun
10. \bar{n} to complement of the copula
11. a to adjective
12. p_2 to past participle
13. i to infinitive

We can assign a separate type s_4 to the imperative tense, but prefer not to, since it is a special case of the present and sometimes of the subjunctive tense. We postulate the following rules

$$\pi_k \rightarrow \pi, \quad s_j \rightarrow s, \quad q_j \rightarrow q, \quad t_l \rightarrow t$$

where π, s, q, t are also basic types and stand for their corresponding generic types, that is when the person or tense does not matter. Other rules are

$$n \rightarrow \pi, \quad p \rightarrow \pi, \quad n \rightarrow \bar{n}, \quad a \rightarrow \bar{n}, \quad s \rightarrow \sigma$$

The simple declarative Persian sentence with a transitive verb has the following structure

subject + object + prepositional phrase + transitive verb.

For example, the following is the Persian sentence for ‘he bought the book from the bookshop’

u ketab ra az ketabkhaneh kharid.
He the book from the bookshop bought.

In this sentence, ‘u’ is the subject, ‘ketab ra’ is the direct object, ‘az ketabkhaneh’ is the prepositional phrase, and ‘kharid’ is the transitive verb in simple past tense. Direct object is followed by the post-position ‘ra’, and the prepositional phrase is headed by a preposition such as ‘az’, ‘dar’, ‘be’, etc. Similar to the English sentence, the prepositional phrase is optional, also (and different from English) since the verb gets fully conjugated, in case of no ambiguity the subject can be omitted³. We assign the compound type $((w^r)o^r(\pi^r)s)$ to the transitive verb, where the inner brackets indicate the optional types and the outer brackets are the usual notation for compound types⁴. This typing tells us that a transitive verb forms a sentence s that might need a (more precisely, an external) subject π , then a prepositional phrase w and finally a direct object o . Similarly, the intransitive verb will have the compound type $((w^r)(o^r)(\pi^r)s)$. These types lead to the following reduction for the sentence with a transitive verb

subject object pp verb \rightarrow statement
 π o w $((w^r)o^r(\pi^r)s)$ \rightarrow s.

The direct object consists of a noun or a pronoun followed by the post-position ‘ra’. In our sample sentence, ‘ketab’ is a noun and is followed by post-position ‘ra’. It is to the combination ‘ketab ra’ that we assigned the type o . To analyze the type of each element of this combination, we assign the type $(\pi^r o)$ to the post-position ‘ra’ and n to ‘ketab’. We use the rule $n \rightarrow \pi$ and get the following reduction for our sample direct object

ketab ra \rightarrow object
 n $(\pi^r o)$ \rightarrow o

³In this case, the personal suffixes that attach themselves to the end of the verb act like the subject. Analyzing the structure of the Persian verb with all of its prefixes and postfixes is left to future work.

⁴These options can also be stated as a meta-rule that each transitive verb of the type $((w^r)o^r(\pi^r)s)$ has also types $((w^r)o^r s)$, $((o^r\pi^r)s)$, and $((o^r)s)$.

The prepositional phrase consists of a noun or a pronoun and is headed by a preposition. Similar to the direct object case, the combination of the noun and the preposition will have type w . We assign the type $(w\pi^l)$ to the preposition. As an example, our sample prepositional phrase ‘az ketabkhaneh’ with ‘ketabkhaneh’ the noun and ‘az’ the preposition, reduces as follows

$$\begin{array}{ccccccc} \text{az} & \text{ketabkhaneh} & \rightarrow & \text{prepositional phrase} \\ (w\pi^l) & n & \rightarrow & w \end{array}$$

An example of a pronoun in the prepositional phrase is the following phrase, which means ‘from him’

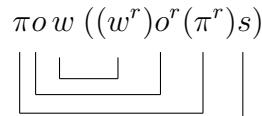
$$\begin{array}{ccccccc} \text{az} & \text{u} & \rightarrow & \text{prepositional phrase} \\ (w\pi^l) & \pi & \rightarrow & w \end{array}$$

The full typing of our sample sentence, with the internal structure of object and prepositional phrase, is as follows

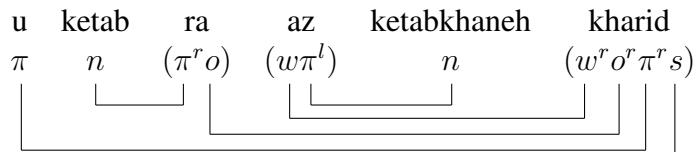
$$\begin{array}{ccccccccccc} \text{u} & \text{ketab} & \text{ra} & \text{az} & \text{ketabkhaneh} & \text{kharid} & \rightarrow & \text{statement} \\ \pi & n & (\pi^r o) & (w\pi^l) & n & (w^r o^r \pi^r s) & \rightarrow & s \end{array}$$

4 Reduction Diagram

We demonstrate the pattern of our reductions, that is the order of application of contraction inequalities, by assigning a vertical line to each basic type and connecting the reduced adjoint types via horizontal lines. The horizontal lines correspond to the bottom part (or contraction part) of the diagrams of [23] for bi-categories, also to the ϵ -triangles of [1, 12] for Compact Closed categories used for quantum protocols. For example, the reduction diagram of a Persian transitive sentence with external subject and object is as follows

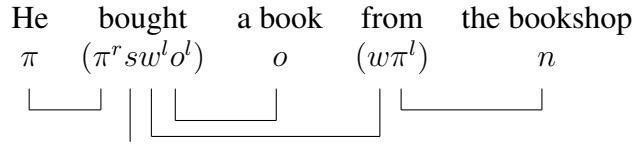


The diagram of the full typing of our sample sentence will be as follows

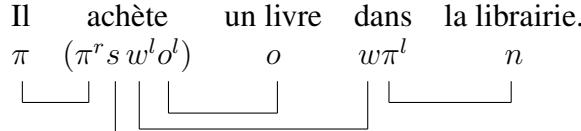


In a grammatical sentence, after the reduction, there should exist only one vertical line for the desired type, for example s in the above case. These diagrams are useful in comparing the reduction pattern of different languages, we provide some examples⁵. A simple English sentence with a transitive verb has the following reduction pattern

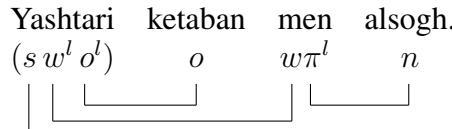
⁵The analysis of English, French, and Arabic sentences are from [15, 3, 2]



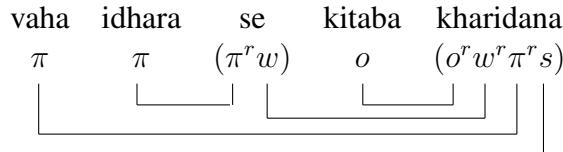
The same sentence in French has the same structure⁶



We observe a different pattern in Arabic, especially in the common form when the subject is omitted



It turns out⁷ that the reduction pattern of the same sentence in Hebrew (with an omitted subject) is similar to that of Arabic. Also Hindi and Persian have similar reduction patterns. For example, the following is the reduction pattern of the sentence 'he bought the book from here' in Hindi



Degree of Nested Contractions. We introduce two numbers for the reduction pattern of each sentence. The first one is the number of times a contraction map $xx^r \rightarrow 1$ or $x^l x \rightarrow 1$ is applied to reduce the sentence to its principal type. This number for our sample simple Persian sentence with transitive verb and prepositional phrase is 5 and is determined by counting the horizontal lines of the reduction diagram. The second number is the number of nested contraction maps, which is always less than or equal to the number of contraction maps in a sentence. A sentence might have different numbers of nested contractions. In this case the maximum of them will be the degree of nesting of the whole sentence. In our sample Persian sentence, we have two blocks of 2 and 4 nesting, so the degree of nested contractions is $\max\{2, 4\} = 4$. The second degree is connected to the *chunks of information* discussed in [15], that is, the number of unprocessed tokens in mind while parsing a sentence. These degrees provide us with a numerical way of comparing the reduction patterns of different languages. In our examples the simple sentences in English and French have 4 contractions, and their degree of nesting is only 2, where as in Arabic these numbers decrease to 3 and 2.

⁶More precisely, 'dans' has the type $(w\tilde{n}^l)$ where \tilde{n} is the type of the noun phrase, for the details see [3].

⁷That is, from discussions with people who know these languages

5 Typing Questions and Compound Sentences

Two forms of questions can be built from a declarative sentence: yes-no questions and wh questions. The yes-no question is formed by placing the question word ‘aya’ (whether) in front of a declarative sentence without changing the order of the words. For our sample sentence, the question ‘Whether he bought the book from the bookshop?’ is as follows

aya	u	ketab	ra	az	ketabkhaneh	kharid?	→	question
(qs ^l)	π	o	w	(w ^r o ^r π ^r s)		→	q	

The wh or ‘che’ questions are more complicated. They enable us to form questions about different parts of the sentence. These questions are built by placing a question word starting with ‘che’ in the place of the part of the sentence about which the question is being asked. For example, the question ‘who bought the book from the bookshop?’ is about the subject and is formed by placing the question word ‘che kasi’ for the subject of our sample sentence, resulting in the following reduction

che	kasi	ketab	ra	az	ketabkhaneh	kharid?	→	question
(q s ^l π)	n	(π ^r o)	(wπ ^l)	n	(w ^r o ^r π ^r s)			

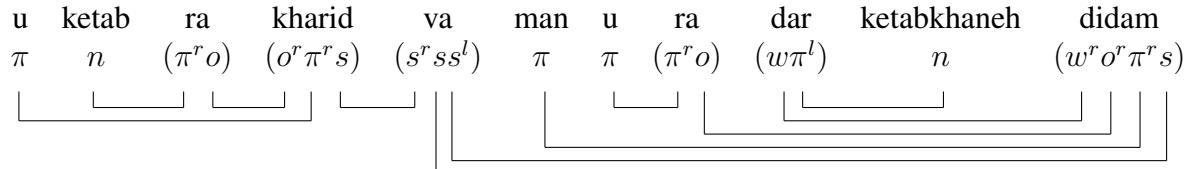
For the wh-question about the direct object of a declarative sentence, we assign the type $(\pi^r qs^l \pi o)$ to the question word ‘che chizi ra’. Since ‘che-chizi-ra’ comes between the subject and the verb, it contains both a π^r to match the subject and a π to match the π^r of the verb. For example, the question ‘what did he buy from the bookshop’ reduces as follows

u	che	chizi	ra	az	ketabkhaneh	kharid?	→	question
π	(π ^r qs ^l πn)	(π ^r o)	(wπ ^l)	n	(w ^r o ^r π ^r s)		→	q

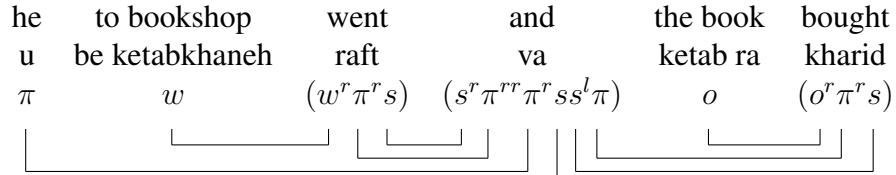
The question about the prepositional phrase is typed similarly. For instance, to form the question ‘where from did he buy the book’ , ‘az koja’ places the prepositional phrase and the reduction is as follows:

u	ketab	ra	az	koja	kharid?	→	question
π	o	(o ^r π ^r qs ^l πow)	(w ^r o ^r π ^r s)			→	q

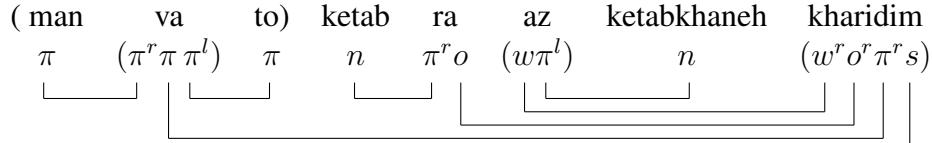
Coordinate Sentences. The simplest compound sentence is formed by putting a conjunction word such as ‘va, amma, ya’ (and, but, or) between two propositions. These conjunction words have the type $(s^r ss^l)$ when they connect two propositions and a generic type of $(x^r xx^l)$ when they connect different constituents of the sentence. For example, the sentence ‘he bought the book and I saw him in the bookshop’ types as follows (‘man’ means ‘I’, ‘dar’ means ‘in’, ‘didam’ means ‘saw’ conjugated for first person singular)



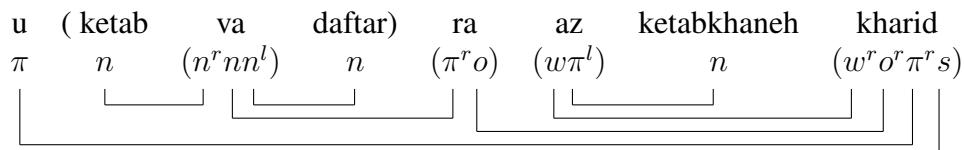
When the subjects of the two conjuncts are the same, the subject will be omitted from the second one. An example is the sentence 'he went to the bookshop and bought the book', with a reduction as follows



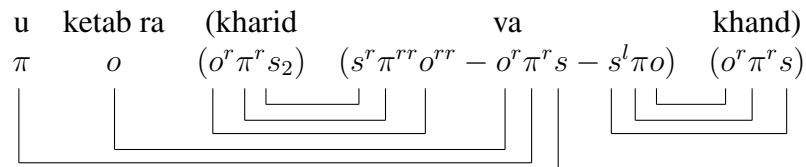
Note that in this case the type of the conjunction 'va' follows the pattern $x^r xx^l$ with $x = \pi^r s$. In sentences with multiple subjects and objects, the conjunction respectively gets the types $(\pi^r \pi \pi^l)$ and $(o^r o o^l)$. For example, the sentence 'I and you bought the book from the bookshop' reduces as follows ('to' means 'you' and 'kharidim' is the first person plural conjugated form of 'bought')



For an example on multiple objects, consider the sentence 'he bought a book and a notebook from the bookshop', which types as follows:

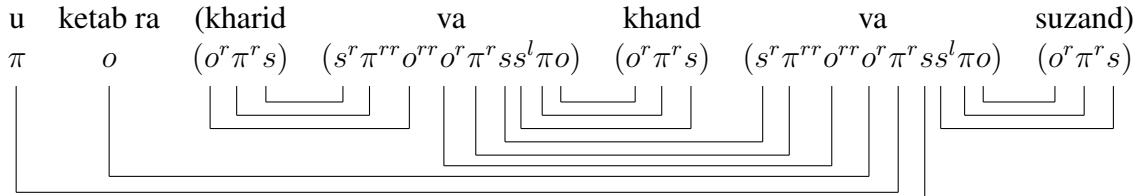


In sentences with multiple verbs, 'va' gets a more complicated type but stays of the form $(x^r xx^l)$. For instance in the sentence 'he bought and read the book' we have $x = o^r \pi^r s$ and the reduction is as follows



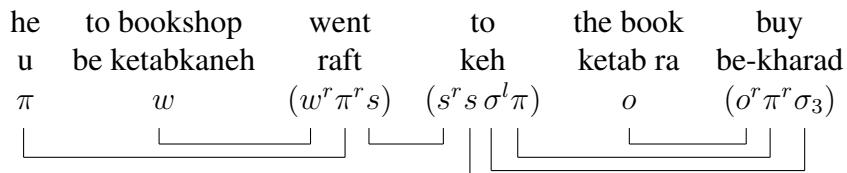
Now the degree of nesting is 5, but if we add the prepositional phrase 'az katabkhaneh' after the object, x becomes of type $w^r o^r \pi^r s$ and the number of nested contractions increases to 7. The increase of complexity of the sentence is at the same time reflected in the iterated adjoints π^{rr} , o^{rr} and w^{rr} in the type of 'va'. This example also demonstrates the importance of the order

of application of adjunctions in the correct reduction of the sentence. For example the π and o from 'u' and 'ketab ra' should not be composed with the π^r and o^r of the 'kharid'. They should wait till 'kharid' composes with $s^r\pi^{rr}o^{rr}$ of 'va' and 'khand' composes with $s^l\pi o$ of 'va' and then compose with $o^r\pi^r$ of 'va'. If we increase the number of verbs, the number of nestings will increase but the number of iterated adjoints of 'va' will stay the same. For example the sentence 'he bought and read and burnt the book' with three verbs types as follows

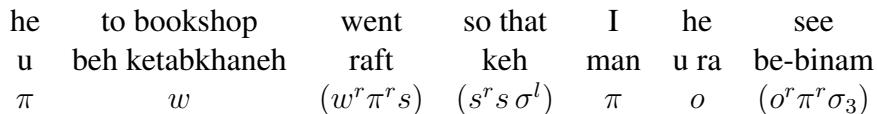


Subordinate Sentences. In these compound sentences, the action of the subordinate clause is dependent on the action of the main clause. The dependency can be causal, to express the purpose of an event⁸, or temporal and each case requires a different analysis. In the causal case, the main clause is written first and its verb is followed by conjunctions such as 'keh' and 'ta'. The subordinate clause follows the conjunction and its verb is either in the subjunctive or copula form. The subjunctive form of the verb is a sentential complement. We use the basic type σ_i for the sentence complement in i 'th tense with the reduction rules $\sigma_i \rightarrow \sigma$ and $s \rightarrow \sigma$. Like in English, when the two clauses share the same subject, the subject of subordinate clause will be omitted. A good way to distinguish this form of dependency from the temporal case is to ask a 'why' question about the verb of the main clause. The compound sentence will then serve as an answer to this question.

For example, the sentence 'u be ketabkaneh raft keh ketab ra bekharad' has a subjunctive subordinate verb ('be-kharad') and is the translation of 'he went to the bookshop to buy the book'. This is a causal example and an answer to the 'why' question 'chera u be ketabkhaneh raft?', translated to 'why did he go to the bookshop?'. We assign the type $(s^r s \sigma^l \pi)$ to the conjunction, which is of the form $x^r xy$ and type our sample sentence as follows



Another causal example when the two clauses do not share the same subject is the sentence 'he went to the bookshop so that I see him'. In this case, the conjunction takes a simpler type $(s^r s \sigma^l)$ and the typing is as follows



⁸I was reminded by J. Lambek that this is what Aristotle calls the "final cause".



An instance of the temporal dependency is when the main clause happens after the subordinate clause. In this case, the subordinate clause is written first and the main clause comes afterwards. The conjunction 'keh' is placed after either the object or the prepositional phrase of the subordinate clause. The whole sentence serves as the answer to the 'when' question about the main clause. The two clauses can share the same subject (in which case it is omitted from the main clause) and no special form is required for the subordinate verb. In particular, when the verb of the main clause is in present continuous or future tenses, the subordinate verb will be in subjunctive tense⁹. The conjunction gets a new form $x^r yx$, reflecting the change of order of the main and subordinate clauses in comparison to the causal case where we had $x^r xy$. Here we have $y = s\sigma^l \pi \sigma^l$ and the two σ 's allow for the subjunctive tenses in both clauses. Moreover we have $x = \pi x'$ where x' is the type of the part of the subordinate sentence that comes before 'keh'.

An example is the sentence 'u ketab ra keh kharid, be khaneh raft', which is the translation of 'after he bought the book, he went home'. The verb of the main clause is 'raft' and happens after the verb of the subordinate clause 'kharid'. The conjunction 'keh' comes after the object of the subordinate clause, so we have $x' = o$, and the two clauses share the same subject. This example is the answer to the question 'u che-vaght be khaneh raft?', which translates to 'when did he go home?'. The typing is as follows

he	the book	after	bought	to home	went
u	ketab ra	keh	kharid,	be khaneh	raft
π	o	$(o^r \pi^r s \sigma^l \pi \sigma^l \pi o)$	$(o^r \pi^r s)$	w	$(w^r \pi^r s)$

An example for when 'keh' follows the prepositional phrase of the subordinate clause is the sentence 'u be ketabkhaneh keh raft, ketab ra kharid', which means 'after he went to the bookshop, he bought the book'. For this sentence we have $x' = w$ and the following typing

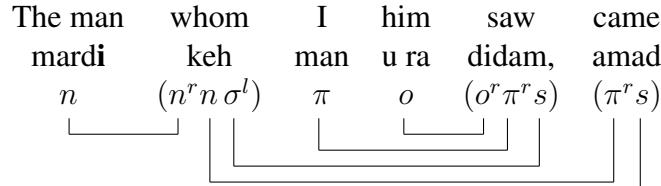
he	to the bookshop	after	went	the book	bought
u	be ketabkhaneh	keh	raft,	ketab ra	kharid
π	w	$(w^r \pi^r s \sigma^l \pi \sigma^l \pi w)$	$(w^r \pi^r s)$	o	$(o^r \pi^r s)$

Relative Sentences. The closest a compound sentence gets to the English relative clause is when more information is provided about the subject, object, or prepositional phrase of a sentence. In each case, the part in focus is usually suffixed by the indefinite morpheme 'i', then followed by the conjunction 'keh', and finally explained by the extra information in the relative clause. The extra information is placed directly after 'keh', which plays the role of a

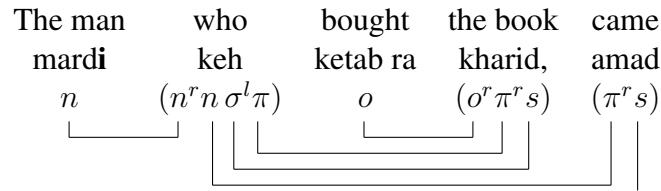
⁹An example is the sentence 'u ketab ra keh be-kharad, be khaneh miravad', which means 'after he buys the book, he will go home'. An example for the subjunctive tense in the main clause is the sentence 'u bayad ketab ra keh kharid, be khaneh be-ravad', which means 'after he buys the book, he should go home'.

relative pronoun. In this form, the whole sentence serves as the answer to the 'che' question about the part that is being explained. The generic type of the conjunction 'keh' will be of the form $(n^r ny)$, same as the $(x^r xy)$ of the causal subordinate sentences. Here, variable y takes different instantiations depending on the parts of the sentence that is being explained, but in all cases it starts with σ^l . The reason for choosing σ over s is that the verb of the relative clause can also be in subjunctive tense¹⁰.

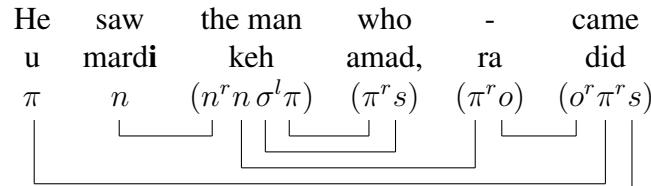
The simplest example is the 'whom' relative clause about the subject, for example in the sentence 'the man whom I saw, came', which types as follows



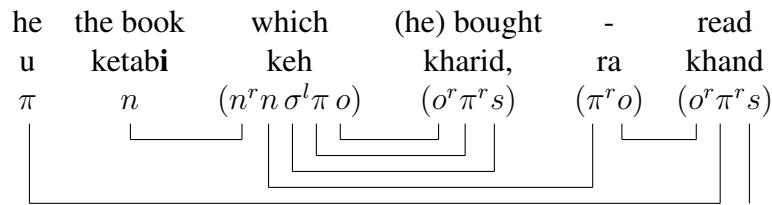
An example of the 'who' relative clause about the subject is the sentence 'the man who bought the book, came', which types as follows



Both of these sentences are answers to the 'che' question 'che mardi amad?', which translates to 'which man came?'. An example of the 'who' relative clause about a person object is the sentence 'he saw the man who came', which types as follows



This sentence is the answer to 'che' question 'u che mardi ra did?', which translates to 'which man did he see?'. Finally, an example of the 'which' or 'that' relative clause about a non-person object is the sentence 'he read the book that he bought', which types as follows



¹⁰Examples are 'mardi keh man u ra be-binam, mimirad', which means 'the man whom I see shall die', 'mardi keh ketab ra be-kharad, mimirad', which means 'the man who buys the book shall die', 'u mardi keh be-ayad ra mikoshad', which means 'he shall kill the man who comes', and 'u ketabi keh be-kharad ra mikhanad', which means 'he shall read the book that he buys'.

Indirect Sentences. Verbs such as 'goftan, danestan, fekr-kardan' (say, know, think) etc, require a sentential complement. We analyze the verb 'goftan' , others are done similarly. The direct and indirect forms of speech are not distinguished in Persian and it is not required that the tense or the person of the second clause changes. For example 'he said that he went to the bookshop' is expressed as 'he said I/he go/goes to the bookshop' and is typed as follows

u	goft	man/u	be	ketabkaneh	mi-ravam/mi-ravad
π	$(\pi^r s \sigma^l)$	π	w	$(w^r \pi^r s)$	

Degree of nestings is 4, but if the verb of the second proposition is transitive, it increases to 5. Sometimes the conjunction 'keh' is used to link the two clauses where it takes a new meaning and thus a new type

u	goft	keh	man	be	ketabkaneh	mi-ravam
π	$(\pi^r s \sigma^l)$	$(\sigma \sigma^l)$	π	w	$(w^r \pi^r s)$	

We can form different questions about this sentence, for example the 'aya' question is as follows

aya	u	goft	man	be	ketabkaneh	mi-ravam?
(qs^l)	π	$(\pi^r s \sigma^l)$	π	w	$(w^r \pi^r s)$	

The question about the subject of the main sentence is as follows

che-kasi	goft	man	be	ketabkaneh	mi-ravam?
$(qs^l \pi)$	$(\pi^r s \sigma^l)$	π	w	$(w^r \pi^r s)$	

We can also have indirect compound sentences, for example the indirect subordinate sentence 'he said that he goes to the bookshop to buy the book' types as follows

u	goft	man	be	ketabkaneh	mi-ravam	keh	ketab ra	be-kharam
π	$(\pi^r s \sigma^l)$	π	w	$(w^r \pi^r s)$	$(s^r s \sigma^l)$	o	$(o^r \sigma_3)$	

6 Poetic Examples

The indirect statement with the verb 'goftan' (to say) is popular with Persian poets. Here we provide an analysis of one verse of the Rubaiyat of Omar Khayyam. The Fitzgerald translation of the poem

"How sweet is mortal Sovrancy!"—think some:
 Others—"How blest the Paradise to come!"
 Ah, take the Cash in hand and waive the Rest;
 Oh, the brave Music of a distant Drum!

The Poem is written as follows in Persian script

گویند بهشت با حور خوش است

من می گویم که آب انگور خوش است

این نقد بگیر و دست از آن نسیه بدار

کواز دهل شنیدن از دور خوش است

We type the first verse as follows (note that the subject of the first sentence is omitted)

say	paradise	with angels	good	is
guyand	behesht	ba hur	khosh	ast
($s\sigma^l$)	n	w	a	($\bar{n}^r w^r \pi^r s$)

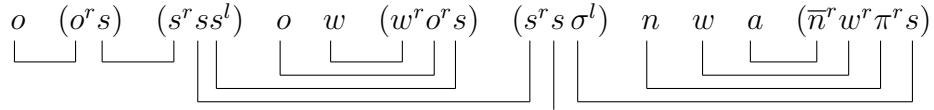
I	say	that	grape juice	good	is
man	miguyam	keh	ab-e angur	khosh	ast
π	($\pi^r s\sigma^r$)	($\sigma\sigma^l$)	n	a	($\bar{n}^r \pi^r s$)

The first sentence has an occurrence of the verb 'astan' (to be) in its non auxiliary copula role. Like in English, the copula needs a subject and a complement (predicate) and also sometimes a prepositional phrase. We assign the type \bar{n} to the complement of the copula with the reduction rule $n \rightarrow \bar{n}$. Thus our 'astan' copula gets the type $(\bar{n}^r (w^r) \pi^r s)$. One of our complements is an adjective, to which we assign the new type a with the reduction rules $a \rightarrow \bar{n}$.

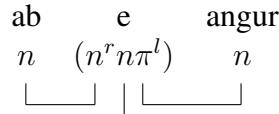
The two parts of the second verse form a subordinate compound sentence where the subordinate sentence of the first part contains two sentences linked by the conjunction 'o', the short form of 'va'. The main clause of the subordinate sentence has an omitted subject. One reduction option is to think of the verbs 'bedar' and 'shenidan' as simple transitive verbs where their post position 'ra' is omitted for the rhyme. We add these omitted 'ra's to the objects 'dast' and 'avaz-e dohol' and obtain the following reduction

this cash	take	and	hand	from that offer	keep
in naghd (ra)	begir	o	dast (ra)	az an nesieh	bedar
o	$(o^r s)$	$(s^r ss^l)$	o	w	$(w^r o^r s)$
since	sound of drums hearing	from far	good	is	
keh	avaz-e dohol (ra) shenidan	az dur	khosh	ast	
$(s^r s \sigma^l)$	n	w	a	$(\bar{n}^r w^r \pi^r s)$	

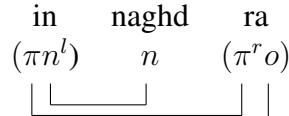
The full reduction of the second verse is as follows and its degree of nesting is 5 (with the internal structure of w , not shown here)



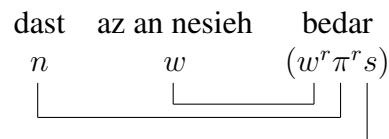
The phrase 'ab-e angur' (grape juice, i.e. wine) is a compound noun phrase formed from two nouns 'ab' (juice) and 'angur' (grape), linked together with the 'ezafeh' phrase 'e'. This is the morpheme that plays a similar role to 'of' or 'from' in English. If we consider the type of the noun phrase the same as noun n then this phrase types as follows



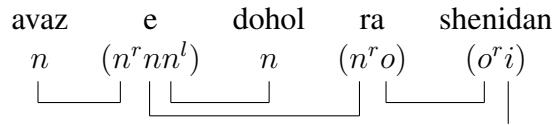
The words 'in' (this) and 'an' (that) in the noun phrases 'in naghd' and 'an nesieh' are the demonstrative pronouns. We use the type πn^l for these pronouns and for example type 'in naghd ra' as follows



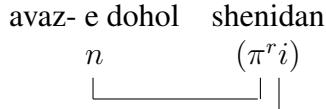
We have treated the word 'dast ra' in the third part as an object for the transitive verb 'bedar'. Another option would be to merge the object with the verb and form a non-transitive *compound verb* 'dast bedar' (refrain). This is a common way of forming compound verbs in Persian. In our poem, and again a common phenomenon in compound Persian verbs, the verb complement 'dast' has been separated from the main part 'bedar' by the prepositional phrase 'az an nesieh'. The reduction for this option is as follows



Finally, in the fourth part of the poem we have another interesting noun phrase 'avaz-e dohol shenidan'. This phrase consists of the complete infinitive 'shenid-an' (the dictionary form of the verb 'shenid' that means to hear), for which we use the type i with the reduction rule $i \rightarrow n$. If we think of 'shenidan' as a simple transitive verb we have to add the omitted post-position 'ra' to the 'ezafeh' phrase 'avaz-e dohol' to get the following typing



However, and similar to the case of 'dast bedar', instead of assuming that the subject and the post position 'ra' have been omitted, one can also think of the phrase 'avaz-e dohol shenidan' as a non-transitive compound verb and get the following reduction



Next we analyze the first two verses of a Ghazal from Khwaja Shams ad-Din Mohammad of Shiraz, known as Hafiz (1310-1325 A.D.)

Goftam gham-e to daram Gofta gham-at sar-ayad
Goftam keh mah-e man sho Gofta agar bar-ayad

گفتم غم تو دارم
گفتا غم سر آید

کفتم که ماه من شو
کفتا اگر برآید

The reduction (also considering the omitted post-positions *ra*) is as follows

(I) said your sorrow have
 goftam gham-e to (ra) daram
 (so^l) o ($o^r s$)

(she) said your sorrow ends
 gofta ghamat sar ayad
 ($s\sigma^l$) n ($\pi^r s$)

(I) said	that	become	my moon
goftam	keh	mah-e man	sho
($s\sigma^l$)	(σs^l)	n	($\bar{n}^r s$)

(she) said	if	ascends
gofta	agar	bar-ayad
$(s\sigma^l)$	$(\sigma\sigma^l)$	(σ_2)

The word 'agar' (if) in the last part is the conditional conjunction used to form conditionals and sometimes (for example in this case) needs to be followed by a subjunctive verb. There are two 'ezafeh' phrases in this verse 'mah-e man' (my moon) and 'gham-e to' (your sorrow), they both stand for possession. So they are possessive phrases, built from a noun 'mah' and 'gham' plus 'e' and the personal pronouns 'man' and 'to'. We type the possessive 'e' as follows

$$\begin{array}{ccc}
 \text{mah} & \text{e} & \text{man} \\
 n & (n^r n \pi^l) & \pi
 \end{array}$$

Another instance of this kind of possessive noun phrase is in the second part of the first verse 'ghama-t', which is the short form of 'gham-e to'.

7 Conclusion and Future Work

We have used pregroups to analyze the sentence structure of the Persian language. Starting from simple sentences with simple tense verbs, we derived the types of the sentence constituents, and most importantly the verb. We used these types to analyze compound sentences and to derive the types of the constituents of these sentences, in particular the conjunction words. Our analysis of one verse of the Rubaiyat of Khayyam and two verses of a poem by Hafiz verified the correctness of the types. All through the paper, we measured the *degree of nesting* equal to the maximum of the number of nested contractions, obtained from the *reduction diagram* of a sentence. In our analyzed examples, the highest degree of nesting is 8 and happens in a compound coordinate sentence with multiple simple tense verbs. This is in the range of 7 ± 2 as the maximum of chunks of unprocessed information that the brain can hold in the short-term memory [15].

Most of our analysis was based on the existence of external subjects and objects in sentences. However, in a sentence with an omitted subject, the personal suffixes that attach to the end of the verb act internally as the subject pronoun. These sentences are typed by assigning the type $\hat{\pi}_k$ to the k 'th person subject pronoun and stating a meta-rule that every verb of the form $(w^r o^r \pi_k^r s)$ also has type $(w^r o^r s \hat{\pi}_k^l)$. We then obtain the following typing for these sentences

$$\begin{array}{ccccccc}
 \text{object} & \text{pp} & \text{verb} & \text{subject pronoun} & \rightarrow & \text{statement} \\
 o & w & (w^r o^r s \hat{\pi}_k^l) & \hat{\pi}_k & \rightarrow & s
 \end{array}$$

We can go further and consider the interesting case when the object is also replaced (in case of no ambiguity) by an object pronoun that is attached to the end of the verb after the subject pronoun. We assign the type \hat{o} to the object pronoun and state a meta-rule, that any verb of the type $(w^r o^r \pi^r s)$ has also the type $(w^r s \hat{o}^l \hat{\pi}_k^l)$. This motivates the following typing:

$$\begin{array}{ccccccc}
 \text{pp} & \text{verb} & \text{subject pronoun} & \text{object pronoun} & \rightarrow & \text{statement} \\
 w & (w^r s \hat{o}^l \hat{\pi}_k^l) & \hat{\pi}_k & \hat{o} & \rightarrow & s
 \end{array}$$

An example is the sentence 'kharid-am-ash' where 'kharid' is the verb, 'am' is the subject pronoun (for the first person singular), and 'ash' is the object pronoun. The analysis of this form of sentence needs an algebraic analysis of the internal structure of Persian verbs, with the different prefixes and suffixes that can attach to it. This has been done in [24]. Using this analysis to study the reduction of Persian sentences with internal subjects and objects and also with verbs in compound tenses constitutes future work.

Acknowledgements

I would like to thank J. Lambek for valuable discussions, comments, and encouragements. I would also like to thank P. Panangaden and S. Abramsky for discussions on examples of Hindi and Hebrew sentences.

References

- [1] S. Abramsky and B. Coecke, 'A Categorical Semantics of Quantum Protocols', in Proceedings of the 19th Annual IEEE Symposium on *Logic in Computer Science*, 2004.
- [2] D. Bargelli and J. Lambek, 'An algebraic approach to Arabic sentence structure', *Linguistic Analysis* **31**, 2001.
- [3] D. Bargelli and J. Lambek, 'An algebraic approach to French sentence structure', In P. de Groote, G. Morrill, and C. Retoré (eds.), *Logical Aspects of Computational Linguistics*, pp. 62-78, Springer-Verlag, 2001.
- [4] M. Barr, 'On Subgroups of The Lambek Pre group', *Theory and Application of Categories* **12**, No. 8, pp. 262-269, 2004.
- [5] W. Buszkowski, 'Lambek grammars based on Pregroups', In P. de Groote, G. Morrill, and C. Retoré (eds.), *Logical Aspects of Computational Linguistics*, pp. 95-109, Springer-Verlag, 2001.
- [6] K. Cardinal, 'An Algebraic Study of Japanese Grammar', Master's Thesis, McGill University, Montreal 2002.
- [7] C. Casadio and J. Lambek, "An algebraic analysis of clitic pronouns in Italian", In P. de Groote, G. Morrill, and C. Retoré (eds.), *Logical Aspects of Computational Linguistics*, pp. 110-124, Springer-Verlag, 2001.
- [8] L.P. Elwell -Sutton, *Elementary Persian Grammar*, Cambridge University Press, 1963.
- [9] E. Fitzgerald (translator), *Rubaiyat of Omar Khayyam*, Kessinger Publishing, 2003.
- [10] M.A. Eslami, 'Datur-e Zaban-e Farsi, tajzieh va tarkib', published by Mashal, 1988 AD (1367 Solar).
- [11] Kh.Sh.M.Hafiz Shirazi, *The Gift*, Daniel Ladinsky (translator), Penguin, 1999.
- [12] G.M. Kelly and M.L. Laplaza, 'Coherence for Compact Closed Categories', *Journal of Pure and Applied Algebra* **19**, pp. 193-213, 1980.
- [13] Omar Khayyam, *Rubaiyat of Omar Khayyam*, St. Martin's Press; Reprint edition 1983.
- [14] J. Lambek, 'Iterated Galois Connections in Arithmetic and Linguistic', in *Galois Connections and Applications*, K. Denecke et al. (eds.), pp. 389-397, 2004.

- [15] J. Lambek, 'A computational algebraic approach to English grammar', *Syntax* **7:2**, pp. 128-147, 2004.
- [16] J. Lambek, 'Type Grammar meets German word order', *Theoretical Linguistics* **26**, pp. 19-30, 2000.
- [17] J. Lambek, 'Type grammar revisited', In A. Lecomte et al. (eds.), *Logical Aspects of Computational Linguistics*, Springer LNAI 1582, pp. 1-27, 1999.
- [18] J. Lambek. 'The mathematics of sentence structure'. *American Mathematics Monthly* **65**, 154–169, 1958.
- [19] A. Kislay, 'Pregroup versus English and Polish grammar', In V.M. Abrusci and C. Casadio (eds.), *New Perspectives in Logic and Formal Linguistics*, pp. 129-154, Bulzoni Editore, 2002.
- [20] P. Khanlari, 'Dastur-e Zaban-e Farsi', published by Boniad-e Farhang-e Iran, 1972 AD (1351 Solar).
- [21] A.K.S. Lambton, *Persian Grammar*, Cambridge University Press, 1984.
- [22] M. Moortgat, 'Categorical type Logics', In J. van Benthem and A. ter Meulen (eds.), *Handbook of Logic and Language*, Elsevier, 1997.
- [23] A. Preller and J. Lambek, 'Free Compact 2-Categories', to appear in *Mathematical Structures in Computer Science*.
- [24] S. Rezaei, 'A Grammar of Persian', Manuscript, April 2002.