

# Acquisition of Nouns Before Verbs: Evidence from Persian

---

Child Language Acquisition

**Masoud Jasbi**

**Dr. Inbal Arnon**

# 1. Introduction:

For the last 30 years, there has been an ongoing debate over which category of words children acquire earlier and what factors account for such pattern of acquisition. It seems to be generally accepted that open class or content words appear earlier in children's speech than function words (Fernald & Marchman, 2006) but whether nouns or more exactly object reference words and proper names are the first emerging words universally and cross-linguistically has been much disputed.

From the very early stages of the debate, two major approaches to this issue were assumed. Firstly, the natural or cognitive approach which considers language acquisition as a corollary of cognitive development; therefore relatively simple and early cognitions such as concrete object references are produced earlier linguistically than harder or more complex ones like relational predicates (e.g. verbs).<sup>1</sup> (Gentner & Boroditsky, 2001) Such an approach necessitates universal and cross-linguistic patterns of acquisition. Secondly, a linguistic relativity approach which considers the acquisition pattern, language specific (or culture specific) and thus not universal. The second approach can be considered as an opposition to the first and in weaker and sometimes later versions asserts that multiple factors including the cognitive one account for acquisition patterns. (Gopnik & Choi, 1995; Tardif, 1996; Tardif, Gelman & Xu 1999)

I will briefly describe how these two approaches have interacted in the past 30 years in the next section; however, what remains true and acknowledged by almost all researchers is that whatever the conclusion, "it still rests on a relatively small sample of the world's languages." (Gentner & Boroditsky, 2009) Taking into account the results, considerations and conclusions of the previous studies on the subject, this paper intends to investigate the pattern of noun vs. verb acquisition in a less well-studied language, namely Persian, using a child language corpus.

Persian is an Indo-European language with an SOV word order spoken mainly in Iran, Afghanistan and Tajikistan. One of the relatively unique features of Persian is its verb inventory. Besides simple verbs, Persian uses compound verbs<sup>2</sup> containing a light verb and a preverbal element which can be a noun, an adjective, an adverb or a prepositional phrase (Lambton, 1967; Mahootian, 1997):

- (1) a. N + LV:  
xær kærdæn<sup>3</sup> (donkey doing) 'to fool'
- b. A + LV:

---

<sup>1</sup> In the cognitive literature of this topic, "easiness" of acquisition or "simplicity" of the object acquired is so closely correlated with "early" acquisition that the terms sometimes appear tautologically as if what is acquired earlier is necessarily simpler/easier and vice versa. But this seems like a rather intuitive hypothesis requiring investigation in its own right as it may not be necessarily the case. In biological development for example, genetic timing is the determining factor while the same intuition of "easiness of task" may seem plausible.

<sup>2</sup> Also known as complex predicates (CP) or predicate verbs (PV)

<sup>3</sup> æn at the end of the verbs is the infinitival marker. Compound verbs formed with light verb *kærdæn* 'to do' are usually causative and have an inchoative counterpart with *shodæn* 'to become'.

	pæhn kærdæn	(wide doing)	‘to widen, to spread’
c.	Adv + LV: bala aværdæn	(up bringing)	‘to vomit’
d.	PP + LV: be bad dadæn	(to wind giving)	‘to waste’

(Follia, Harleyb, & Karimi, 2005)

This paper seeks to find whether Persian speaking children acquire nouns before verbs, and among verbs, which type (simple, noun-compound, adjective-compound , etc) is acquired earlier. I shall explain the method used for this study in section 4 (Data) and the results of the inquiries and further discussions and conclusions in sections 5 and 6 (Results and Discussion) respectively.

## 2. Literature Review:

The idea that concepts referred to by nouns are more basic than concepts referred to by verbs is as old as Aristotle. In *Metaphysics Book Zeta*, he explains that when words such as “to walk” or “to sit” are contemplated, they necessitate conception of an entity (substance or individual) for which the action of walking or sitting can happen. Therefore such words (verbs) are not self subsistent and depend on “existing things” and such existing things are more basic concepts or in fact prerequisites to the conception of predicates. (Alston 1964)

Following the same line of reasoning, Gentner (1982) put forth the Natural Partitioning Hypothesis which asserts that the linguistic distinction between nouns and predicates (e.g. verbs and prepositions) is based on a perceptual distinction between concrete concepts and relational concepts and that nouns are conceptually simpler or more basic than predicates. She suggested that this explains why nouns are more dominant in English children’s early vocabulary and they learn them before predicates like verbs (Macnamara, 1972; Nelson, 1973; Gentner, 1978; Bates, Bretherton, & Snyder 1988) and even why children comprehend nouns better than verbs in their early stages. (Goldin-Meadow, Seligman, & Gelman, 1976). She claimed that it is easier for infants to separate objects from the perceptual-cognitive information of their surroundings and map the stream of language being spoken to these already formed cognitions of objects.(Gentner, 1982)

But if the reason of “noun bias” is indeed cognitive, it should thus be universal and observable in all languages. Gentner (1982) presented data from the speech of German, Japanese, Kaluli, Mandarin Chinese and Turkish children to support the predominance of nouns in early vocabularies corsslinguistically. Many later researches also support this universal cognitive view. In a study of Italian and English children, Caselli, Casadio & Bates (1999) observed the same level of noun bias in both languages and according to Bornstein et al. (2004), the noun dominant pattern holds for Dutch, English, French, Hebrew, Italian, Korean and Spanish too. Furthermore Gentner and Boroditsky supported the Natural

Partitioning hypothesis by data from less well studied languages such as Tzeltal (Gentner & Boroditsky, 2001) and Navajo (Gentner & Boroditsky, 2009).

On the other hand, some linguists argued that noun bias in early language acquisition is due to language specific features which make nouns more salient to children rather than cognitive-perceptual regularities and universals. This argument was supported by studies on so-called “verb friendly” languages like Korean and Chinese in which verb inflection is minimal and verbs are close to nouns in morphological behaviour. (Saxton, 2010) In a comparative study of English and Korean children, Gopnik & Choi (1995) reported that Korean children, unlike English children, used verbs to refer to actions and relations, supposedly as a result of verb frequency in Korean motherese compared to English. In another study, Tardif (1996) investigated spontaneous speech samples of ten 22 months old monolingual Mandarin-speaking children and concluded that a variety of linguistic and sociocultural factors (rather than cognitive ones) influence early word learning. She reported that even with a liberal definition of “noun” (including proper names), no noun bias can be observed in the early vocabulary of Mandarin speakers.

However, the studies mentioned from both sides seem to present a contradiction. Why Chinese and Korean are reported to show noun bias in some studies (Gentner, 1982; Bornstein et al. 2004) and not in some others (Tardif, 1996; Gopnik & Choi, 1995)? The answer to this question lies in a crucial methodological issue mentioned by Tardif, Gelman, and Xu (1999).

There are two prevalent methods of measuring children’s vocabulary; First, “spontaneous speech sampling” in which sessions of children’s interaction, usually with their mothers, are recorded for later transcription and analysis, and second, what is known as “MacArthur checklist” which is a list of words presented to mothers so that they can determine what words their children are already producing. Tardif, Gelman and Xu (1999) used both methods to measure the vocabulary of 20 months old English and Mandarin speaking Children. In spontaneous speech sampling, they created three different contexts: noun-favourable (reading a picture book), verb-favourable (playing with a mechanical toy performing various activities) and neutral (playing with different toys).

While Mandarin speaking children showed considerable noun bias in a noun-favourable context, they produced up to two times more verbs than nouns in the neutral and verb-favourable contexts. Besides, English speaking children showed noun bias in the noun-favourable context, no bias in the neutral context and verb bias in the verb-favourable context. This demonstrates that a researcher could report a noun, verb or no bias for children’s early vocabulary depending on the context of the sampling. Comparing the spontaneous sampling and the MacArthur checklist results, Tardif, Gelman and Xu (1999) concluded that the spontaneous sampling method is highly vulnerable to contextual variability and greatly underestimates children’s vocabularies. The underestimation is even more severe in case of nouns since they are more referentially specific than verbs. (Gentner & Boroditsky, 2001)

Using a MacArthur checklist, both Korean and Chinese showed considerable noun bias.<sup>4</sup> (Au et al., 1994; Tardif, Gelman and Xu, 1999)<sup>5</sup>

An alternative to the aforementioned methods is the rather recent experimental one in which children's ability to learn novel nouns and verbs is tested. Childers & Tomasello (2006) reported that English speaking children learn nouns more rapidly than verbs. A similar conclusion was made for Mandarin, Japanese and English speaking children by Imai et al. (2006). They showed that children manage to acquire and generalize novel noun tokens by the time they are three years old across these languages but verbs were not generalized correctly until the age of five. In fact Mandarin children did not manage to acquire verbs correctly and confused them with nouns until the age of eight. This suggests that the concept of a "verb-friendly" language might need revision.

Of course the noun-verb acquisition debate and the related studies are not as simple, brief and straightforward as presented here. Nevertheless, the recent discussions and studies of noun-verb acquisition seem to generally acknowledge the high influence of cognitive-semantic factor and the truth of the Natural Partitioning hypothesis. However, this does not mean that the influence of other factors such as cultural, socioeconomic or linguistic, are rejected. What is widely accepted is that there seems to be a universal tendency for children to linguistically encode certain concepts such as object references before others such as relational predicates while the actual rate can be further influenced by linguistic, cultural, socioeconomic or contextual factors.

### **3. Research Question:**

In this study I seek to answer two questions:

- I. Is "noun bias" present in Persian speaking children's early vocabulary?
- II. Which Persian verb form appears more dominantly in Persian speaking children's early vocabulary: simple verbs, noun-compound verbs, adjective-compound verbs, adverb-compound verbs or prepositional-compound verbs?

### **4. Data:**

In order to answer the proposed questions, I conduct a corpus search of the CHILDES<sup>6</sup> database. Since the research questions are concerned with noun and verb categories, the chosen dataset for this research should be morphologically tagged. Among the Persian corpora in CHILDES database, only the "Neiloufar Family" dataset has morphological tagging. This dataset is comprised of Lilia and Minu corpora and according to the CHILDES

---

<sup>4</sup> On the other hand this noun bias was not as high as English, showing the verb-friendliness of these languages.

<sup>5</sup> A plausible question might be whether MacArthur checklist overestimates children's vocabulary or favours noun bias. There have been several arguments for and against, but it seems that the checklist is generally considered to be a more accurate estimation of children's vocabulary than the sampling technique.

<sup>6</sup> Child Language Data Exchange System at: <http://chilides.psy.cmu.edu/>

manual, both are spontaneous speech samples recorded in a natural environment in the children's homes in Tehran, without the presence of the researcher. Lilia corpus contains 31 sessions of her speech from when she was 1;11 to when she was 2;10 and Minu was recorded in 103 sessions from the age 4;0 to 5;2.

Since this research focuses on “early” acquisition and due to the fact that noun bias is claimed to be particularly present in the first 200 words (Fernald & Marchman, 2006) and children are reported to have an average vocabulary size of 300 words by the time they are 2 years old (Fenson et al. 1993), it is more plausible to choose the Lilia corpus and specially investigate her earlier sessions from age 1;11 to 2 or a little later.

Of course as mentioned in section 2 (Literature), spontaneous speech samples are prone to be influenced by the context and underestimate children's vocabulary. Therefore, besides the quantitative corpus analysis, we also need to discuss the context of the recording and the possible influences on the data.

Reading through the corpus, it can be understood that the recordings have been carried out in the natural home environment of the child with family members around while the mother and sometimes other family members tried to elicit speech from Lilia. These elicitations are sometimes noun-favourable like “what is this?” and sometimes verb favourable like “what did you do?”. But the general context of Lilia's speech samples seems to be relatively neutral. Nevertheless the corpus is surely a limited and underestimated account of Lilia's actual and overall vocabulary.

## 5. Results:

In the first five sessions recorded from the age 1;11 to 2;1, Lilia produced 204 different nouns<sup>7</sup> ranging from kinship terms such as *mami*<sup>8</sup> “mom” and *dadashi* ‘brother’ to *kæfshduzæk* ‘ladybug’ and *boshghab-pærænde* ‘flying-saucer’ or more abstract notions like *ejaze* ‘permission’. The ten most frequent nouns used are: *mami* (44) – *shir* ‘milk’ (31) – *dadashi* ‘brother’ (15) – *del* ‘stomach/heart’ (10) – *dæst* ‘hand’ (10) – *yeki* ‘one’ (10) – *ab* ‘water’ (9) – *buq* ‘horn’ (7) – *nanai* ‘dance’ (7) and *nobæt* ‘turn’<sup>9</sup> (7). Although most of these 204 nouns are concrete or reference words, there are a few nouns which refer to actions such as *koshti* ‘wrestling’, *masaj* ‘massage’ and *tupbazi* ‘ball playing’ or abstract notions like *jævab* ‘answer’ and *nobæt* ‘turn’.

On the other hand Lilia used 97 verbs from which 43 are simple and 54 compound.<sup>10</sup> It should be noted that the simple verbs counted are mostly different paradigms of the same verb and in fact the number of different simple verbs used is at most 15 while the number of compound verbs is relatively accurate. (making the total number of verbs 69) Among the simple verbs the five most frequent are different paradigms of *budæn* ‘to be’, *dashtæn* ‘to have’, *dadaæn* ‘to give’, *xordæn* ‘to eat’ and *avaerdæn* to bring’.

---

<sup>7</sup> Appendix 1

<sup>8</sup> /a/ in Persian is close to the vowel of *car* in English

<sup>9</sup> As in: It is my turn!

<sup>10</sup> Appendices 2 and 3

From 54 compound verbs used, 34 are noun-compounds, 16 are adjective-compounds and 4 are prepositional-compounds. The ten most frequent compound verbs are *dorost kærdaen/shodaen* ‘to make or correct/ to become right’ (14), *bazi kærdaen* ‘to play’ (13), *jish kærdaen/dashtaen* ‘to urinate/have to urinate’ (10), *peida kærdaen/shodaen* ‘to find/be found’ (7), *gerye kærdaen* ‘to cry’ (7), *toopbazi kærdaen* ‘to play with a ball’ (7), *dæst zædaen* ‘to touch with hands’ (5), *bær dashtaen* ‘to pick up’ (5) *tæmoom kærdaen/shodaen* ‘to finish’ (4), *xoob kærdaen/shodaen* ‘to make/become fine’ (4).

Now regarding the whole corpus, Lilia has used 530 different noun types in all 31 sessions recorded from the age 1;11 to 2;10.<sup>11</sup> *Mami* (420) is by far the most frequent noun used while two other forms of ‘mother’, namely *maman* (31) and *mama* (11), are also commonly used. The ten most frequent, excluding the forms for ‘mother’ are: *dadashi* ‘brother’ (86), *shir* ‘milk’ (55), *xale* ‘auntie’ (31), *ab* ‘water’ (24), *babai* ‘father’ (22), *kar* ‘work’ (19), *doone* ‘grain/unit’<sup>12</sup> (18) *soop* ‘soup’ (14), *yeki* ‘one’ (12) and *dæst* ‘hand’ (11).

However, Lilia has produced 104 simple and 129 compound verbs, making a total of 233 verbs in all 31 sessions.<sup>13</sup> But the number of simple verbs is exaggerated again by counting different conjugation forms and the actual number is 25 making the aggregate 154. Among the simple verbs, different paradigms of *budaen* ‘to be’, *didæn* ‘to see’, *goftæn* ‘to say’, *neveshtaen* ‘to write’, *amædaen* ‘to come’ and *dashtaen* ‘to have’ are the most frequent.

With 90 tokens, noun-compounds are still the most frequent among the compound verbs followed by adjective-compounds (34) and prepositional-compound (5). The ten most frequent are: *bazi kærdaen* ‘to play’ (50), *dorost kærdaen/shodaen* ‘to make or correct/ to become right’ (41), *doost dashtaen* ‘to like’ (23), *bær dashtaen* ‘to pick up’ (21), *jish kærdaen/dashtaen* ‘to urinate/have to urinate’ (20), *tæmoom kærdaen/shodaen* ‘to finish’ (13), *dær amædaen* ‘to come out’ (13), *bærænde shodaen* ‘to win’ (11), *nega kærdaen* ‘to look’ (11), *jæm kærdaen* ‘to gather’ (8).<sup>14</sup>

Age / Category	Noun	Total Verb	Simple Verb	Noun Compound	Adjective Compound	Prepositional Compound
1;11 – 2;1	204	69	15	34	16	4
1;11 – 2;10	530	154	25	90	34	5

Table 1: results from the “Family” corpus search

## 6. Discussion:

Regarding the first research question, the data clearly shows a noun bias whether in Lilia’s earlier vocabulary from the age 1;11 to 2;1 or her overall speech till the age 2;10. The number

<sup>11</sup> Appendix 4

<sup>12</sup> Used for counting

<sup>13</sup> Appendix 5

<sup>14</sup> Adverb-compounds are extremely rare and rather archaic in modern Persian. Therefore it’s not surprising that none is present in Lilia’s speech.

of nouns produced by Lilia is almost three times higher than the number of verbs in her early vocabulary and even more than three times higher in the whole corpus.

Among the verbs, the number of noun-compound verbs is twice as much as adjective-compound verbs or simple verbs in her early vocabulary. By the times Lilia is 2;10 years old, the number of noun-compound verbs has reached almost three times the number of simple or Adjective-compound verbs. In fact noun-compounds are used more than all other forms put together. Therefore, considering the answer to the second research question, noun-compound verbs are clearly the most dominant verb forms followed by adjective-compounds, simple verbs and lastly prepositional compounds .

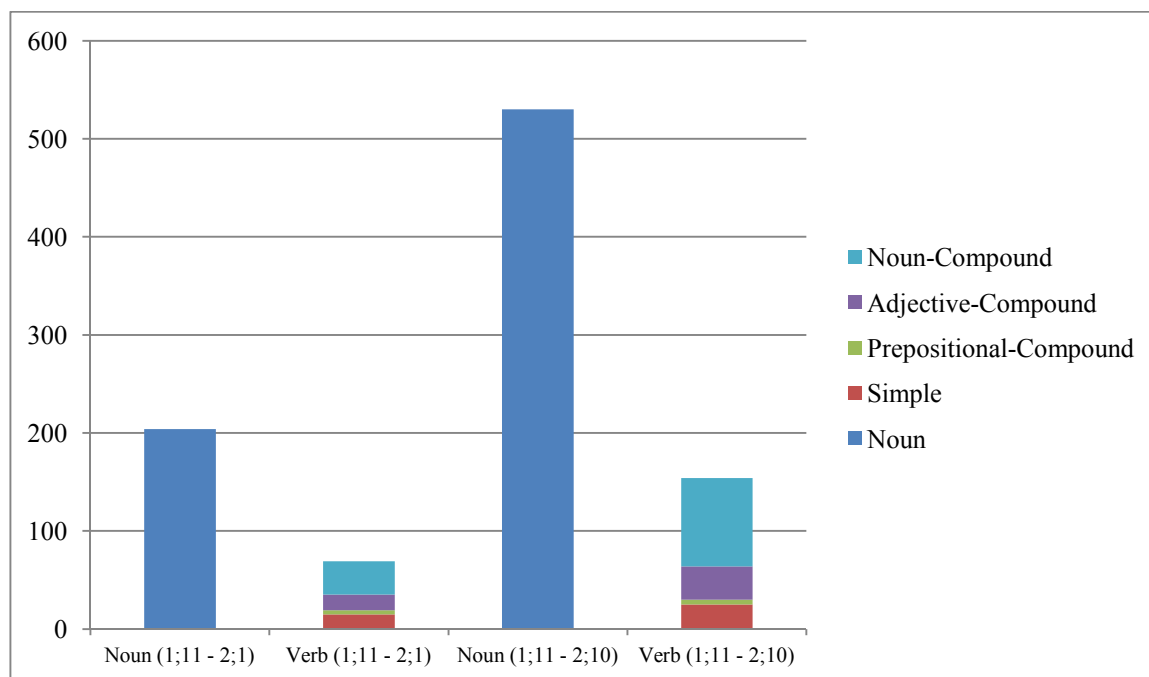


Figure 1: the number of nouns and verbs in Lilia's early (1;11-2;1) and later (1;11-2;10) speech

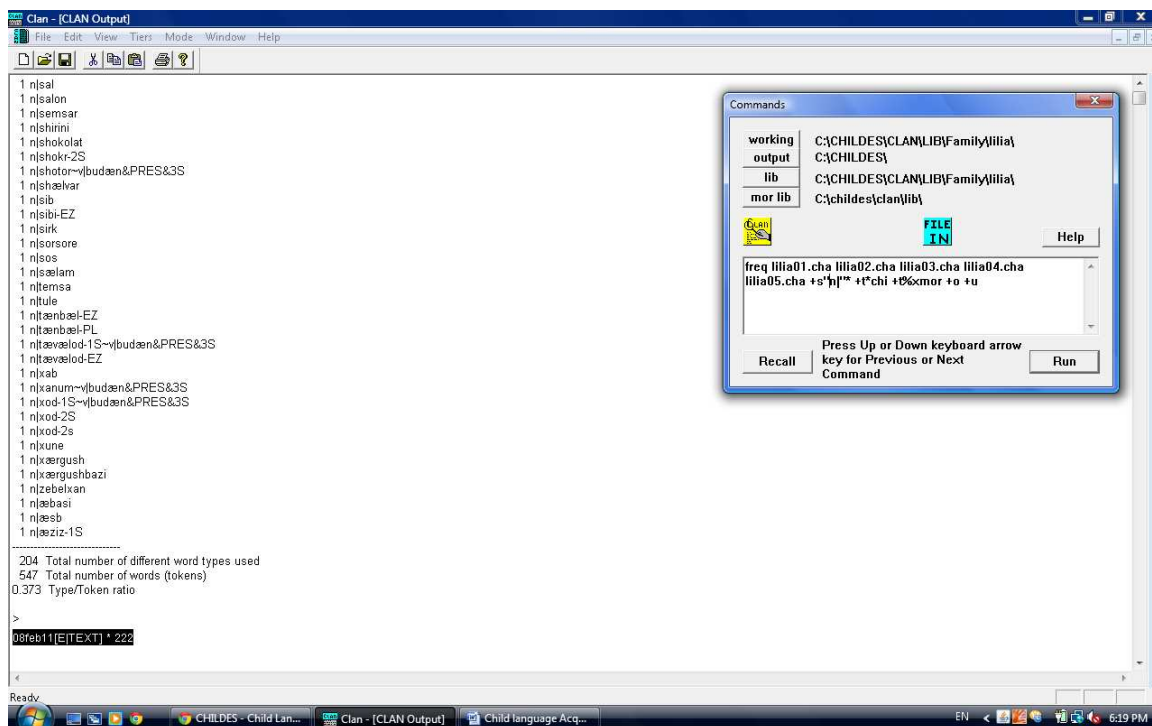
The results obtained from the “Family” corpus evidently confirm the Natural Partitioning Hypothesis. What seems to be novel in these results and more striking than the observed noun bias, is the “noun-compound bias”. While simple verbs are simpler phonologically and morphologically, the results show that Lilia has acquired more noun-compounds and in fact the number of simple verbs is close to the number of adjective compounds. This rules out the phonological or morphological factor and clearly shows the effect of the noun as a preverbal element. Further support for the cognitive view may come from the fact that all the preverbal noun elements produced by Lilia are already present in her speech as independent nouns. If the results of this study are indeed accurate and true regarding the nature of spontaneous speech samples, future studies and further investigations on Persian language acquisition can bring firm evidence for Natural Partitioning Hypothesis and help the development of early word acquisition debates.



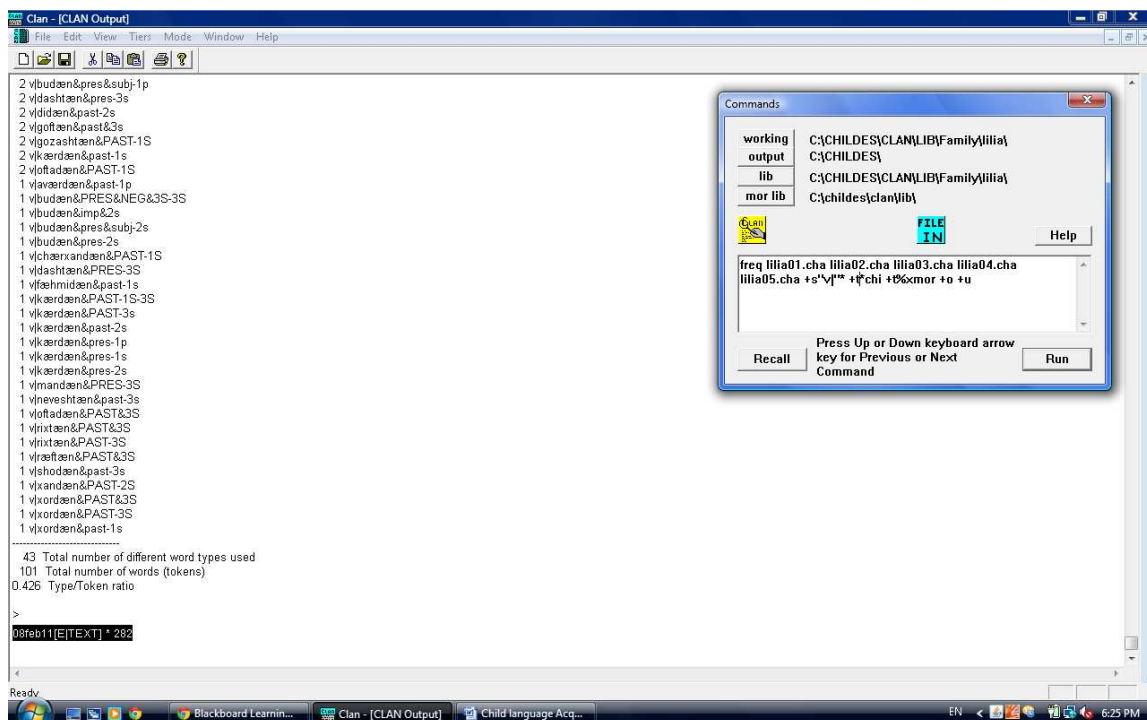
## 7. Appendix:

CLAN Command lines and a screen shot of each output is presented in this section:

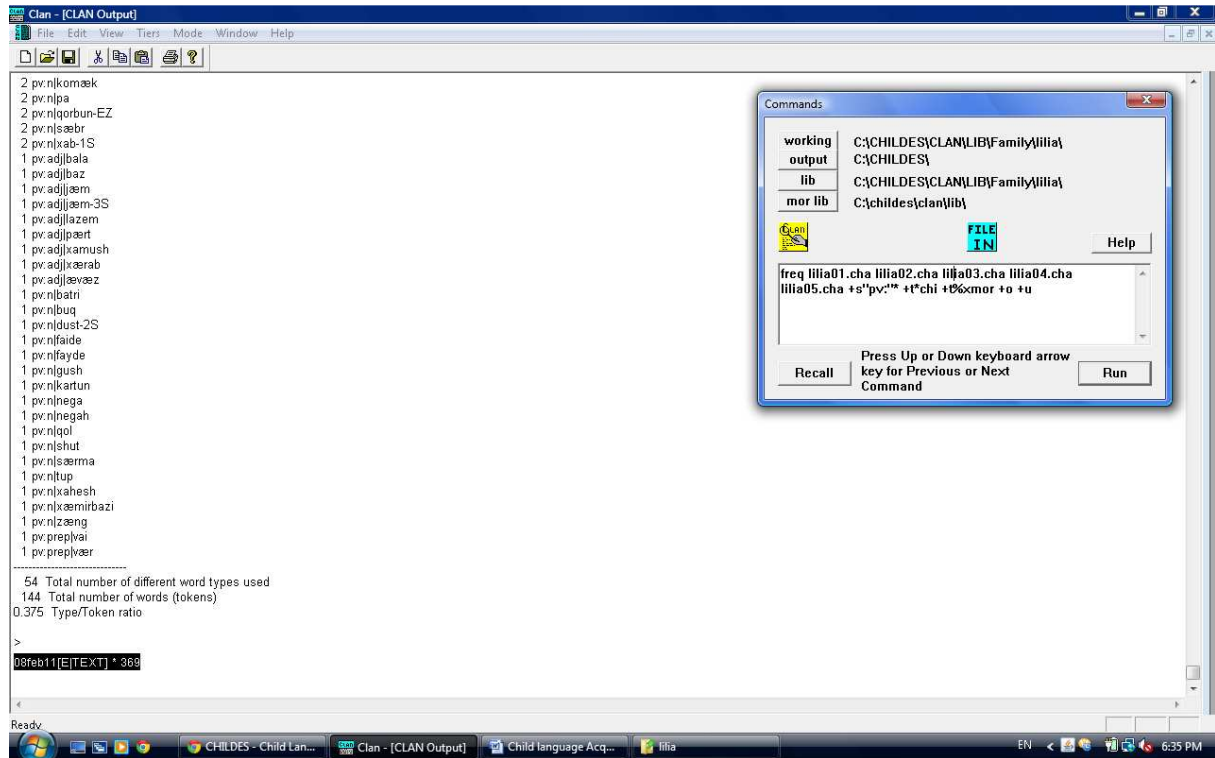
```
1. freq lilia01.cha lilia02.cha lilia03.cha lilia04.cha lilia05.cha
+s"n|"* +t*chi +t%xmor +o +u
```



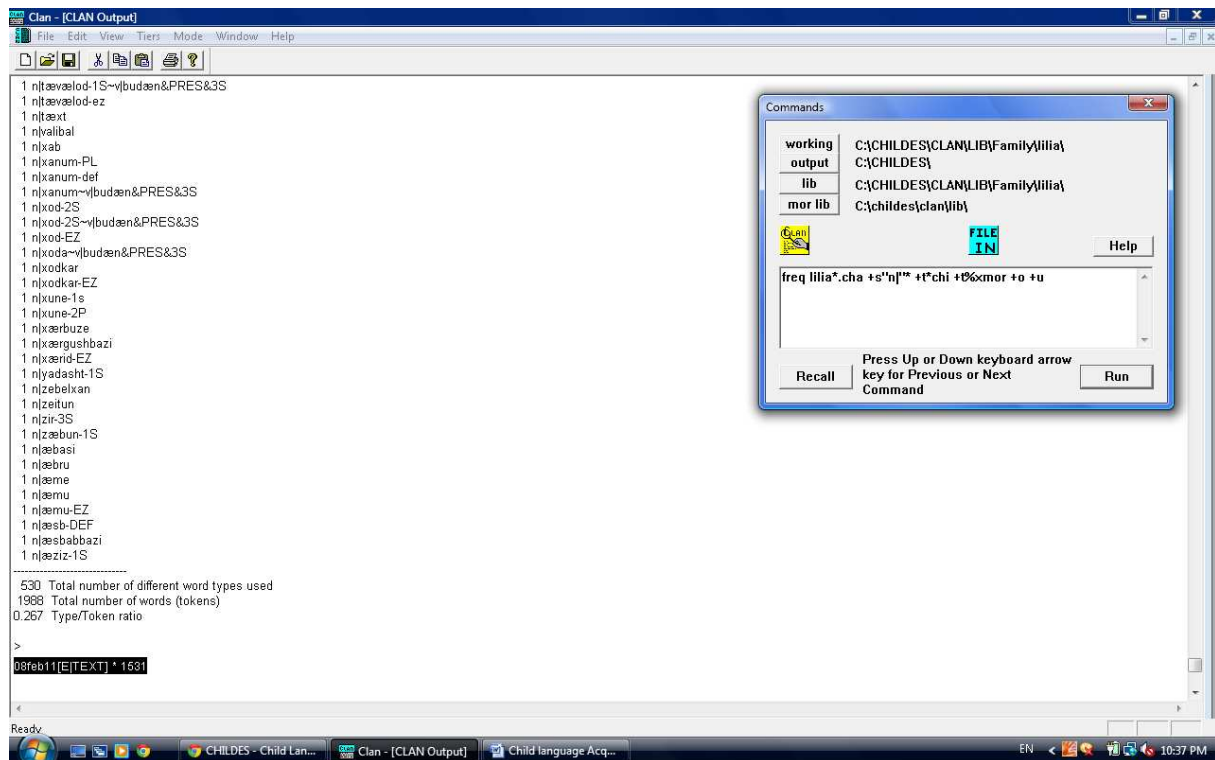
```
2. freq lilia01.cha lilia02.cha lilia03.cha lilia04.cha lilia05.cha
+s"v|"* +t*chi +t%xmor +o +u
```



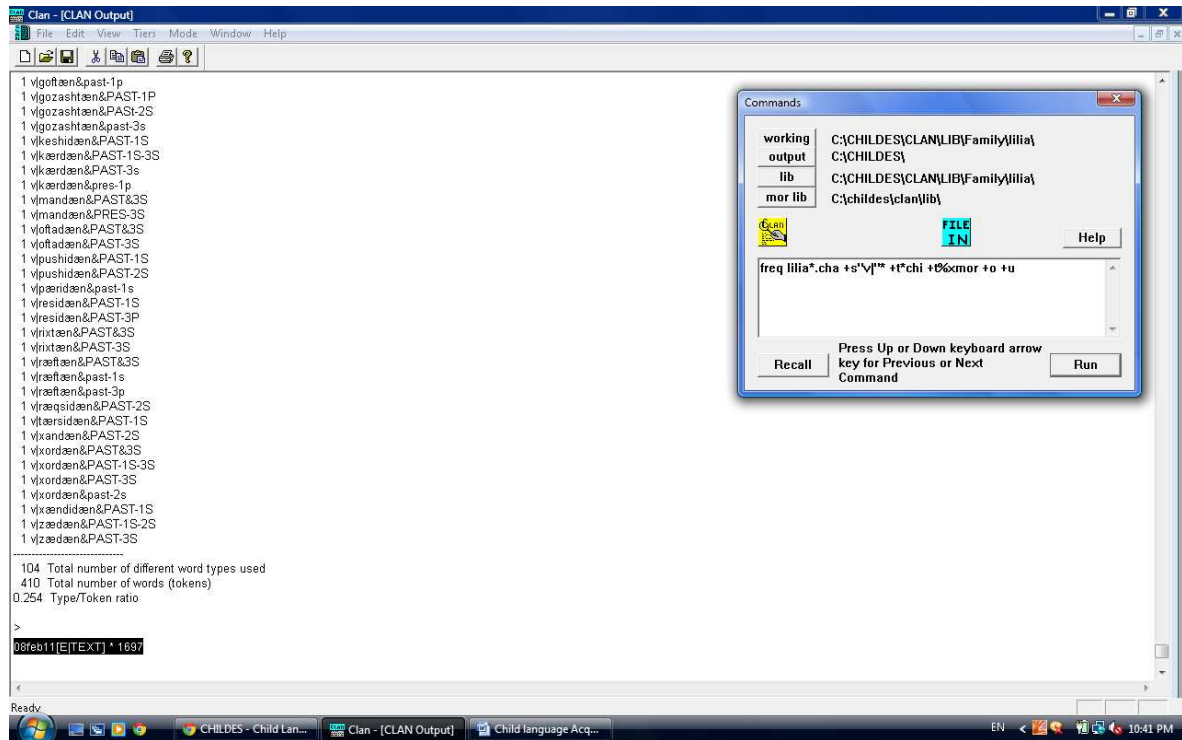
3. `freq lilia01.cha lilia02.cha lilia03.cha lilia04.cha lilia05.cha  
+s"pv:"* +t*chi +t%xmor +o +u`



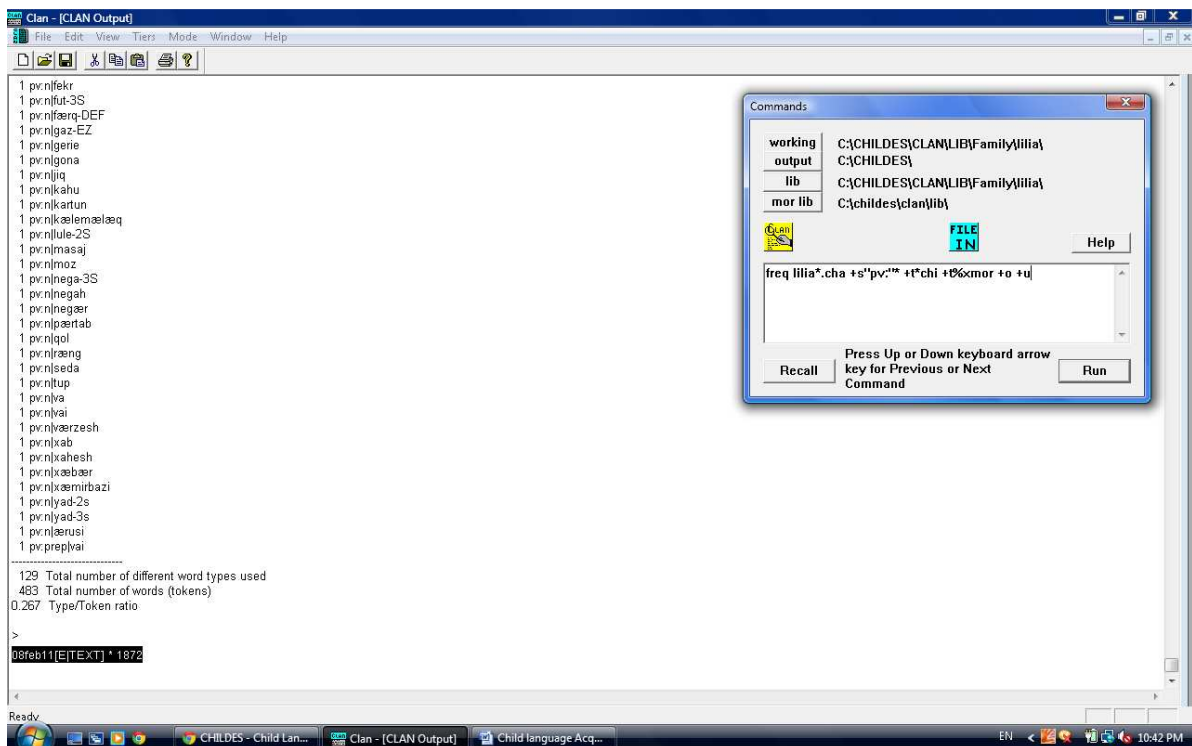
4. `freq lilia*.cha +s"n|"* +t*chi +t%xmor +o +u`



5. freq lilia\*.cha +s"v|"\* +t\*chi +t%xmor +o +u



6. freq lilia\*.cha +s"pv:"\* +t\*chi +t%xmor +o +u



These commands were frequently used to check the context of utterances. Since a satisfactory output could not be shown in one window and the results were not much discussed in the text, only the commands are mentioned here.

7. `combo lilia01.cha lilia02.cha lilia03.cha lilia04.cha lilia05.cha +s"n|"* +t*chi +t%xmor +w2 -w2`
8. `combo lilia01.cha lilia02.cha lilia03.cha lilia04.cha lilia05.cha +s"v|"* +t*chi +t%xmor +w2 -w2`
9. `combo lilia01.cha lilia02.cha lilia03.cha lilia04.cha lilia05.cha +s"pv:"* +t*chi +t%xmor +w2 -w2`
10. `combo lilia*.cha +s"n|"* +t*chi +t%xmor +w2 -w2`
11. `combo lilia*.cha +s"v|"* +t*chi +t%xmor +w2 -w2`
12. `combo lilia*.cha +s"pv:"* +t*chi +t%xmor +w2 -w2`

## 8. Bibliography:

- Alston, W. P. (1964). *Philosophy of language*. Englewood Cliffs, N.J.: Prentice-Hall.
- Au, T. K., Dapretto, M., & Song Y. K. (1994). Input vs. constraints: Early word acquisition in Korean and English. *Journal of Memory and Language*, 33, 567–582.
- Bates, E., Bretherton, I., & Snyder, L. (1988). *From first words to grammar: Individual differences and dissociable mechanisms*. New York: Cambridge University Press.
- Bornstein, M.H., Cote, L.R., Maital, S., Painter, K., Park, S.Y., Pascual, L., Venuti, P. & Vyt, A. (2004). Cross-linguistic analysis of vocabulary in young children: Spanish, Dutch, French, Hebrew, Italian, Korean and American English. *Child Development*, 75(4), 1115-1139.
- Caselli, C., Casadio, P., & Bates, E. (1999). A comparison of the transition from first words to grammar in English and Italian. *Child Language*, 26, 69-111.
- Fernald, A., & Marchman, V. (2006). Language Learning in Infancy. In M. J. Traxler, & M. A. Gensbacher, *Handbook of Psycholinguistics* (pp. 1027-1071). London: Elsevier.
- Folia, R., Harley, H., & Karimi, S. (2005). Determinants of event type in Persian complex. *Lingua*, 115, 1365-1401.
- Gentner, D. (1978). On relational meaning: The acquisition of verb meaning. *Child Development*, 49, 988-998.
- Gentner, D. (1982). Why Nouns are Learned Before Verbs: Linguistic Relativity versus Natural Partitioning. In S. Kuczaj, *Language Development: Language thought and culture* (Vol. 2, pp. 301-334). London: Lawrence Erlbaum.
- Gentner, D., & Boroditsky, L. (2009). Early acquisition of nouns and verbs: evidence from Navajo. In V. Gathercole, *Routes to language: studies in honor of Melissa Bowerman* (pp. 5-36). New York, NY: Taylor & Francis.

Gentner, D., & Boroditsky, L. (2001). Individuation, relativity, and early word learning. In M. Bowerman, & S. Levinson, *Language Acquisition and conceptual development* (pp. 215-256). Cambridge: CUP.

Goldin-Meadow, S., Seligman, M., & Gelman, R. (1976). Language in the two year old. *Cognition* , 4, 189-202.

Gopnik, A., & Choi, S. (1995). Names, relational words and cognitive development in English and Korean Speakers: nouns are not always learned before verbs. In M. Tomasello, & W. Merriman, *Beyond names for things: Young children's acquisition of verbs* (pp. 63-80). Hillsdale, NJ: Erlbaum.

Lambton, A. K. (1967). *Key to Persian Grammar*. Cambridge: Cambridge University Press.

Macnamara, J. (1972). Cognitive basis of language learning in infants. *Psychological Review* , 79, 1-13.

Mahootian, S. (1997). *Persian*. London: Routledge.

Nelson, K. (1973). Structure and strategy in learning to talk. *Monographs of the Society for Research in Child Development* , 38, 149.

Saxton, M. (2010). *Child Language: Acquisition and Development*. Los Angeles: SAGE.

Tardif, T. (1996). Nouns are not always learned before verbs: evidence from Mandarin speakers' early vocabularies. *Developmental Psychology* , 32, 492-504.

Tardif, T., Gelman, S. A., & Xu, F. (1999). Putting the noun bias in context: A comparison of English and Mandarin. *Child Development*, 70(3), 620–635.