Root Infinitives in Child Russian:
A Comparison with Italian and Polish

Eva Bar-Shalom & William Snyder

Department of Linguistics
University of Connecticut
barshalo@uconnvm.uconn.edu, wsnyder@uconnvm.uconn.edu

Abstract
Production of Root Infinitives (RIs) was compared for children acquiring three different languages, two pro-drop languages, Polish and Italian, and a non-pro-drop language, Russian. The greatest differences were found between the Polish and the Russian children. The three Polish subjects produced almost no RIs in the transcripts examined, while the two Russian subjects produced considerable numbers of RIs and did so for a protracted period of time. Among the six Italian children studied, only one produced a substantial number of RIs, and none produced as many RIs as the two Russian children examined. The differences between Russian and Italian are the most pronounced when children are compared by MLU or obtained MLU, and are obscured in an age-based comparison. We propose that languages such as Russian and Polish, which form a near minimal pair with respect to their pro-drop status, are the likeliest to yield clear acquisitional contrasts. Our results are consistent with the Rhee-Wexler hypothesis that the RI phenomenon is associated with non-pro-drop target languages, but larger numbers of children need to be examined to compensate for the considerable individual variation, independent of target language.

1. Introduction
Recent attempts to characterize the languages that do and do not exhibit a Root Infinitive (RI) stage have focused on (a) the richness of verbal morphology (e.g. Wexler 1994, Hoekstra & Hyams 1995), or (b) the possibility of null subjects in the target language (Rhee & Wexler 1995). Either the languages with impoverished verbal morphology, or the languages that disallow null subjects, are proposed as the languages associated with an acquisitional stage of RI's.

Russian provides an important test case for these hypotheses, because it satisfies the proposed criteria for rich verbal morphology, but it does not permit thematic null subjects of the relevant type (cf. Chvany 1973, Franks 1995). Hence, Rhee & Wexler predict that Russian will exhibit an RI stage, while Hoekstra & Hyams predict that it will not.

Several investigators (e.g. Guasti 1993, Phillips 1996) have noted that even children acquiring languages (e.g. Italian, Spanish) that are generally characterized as lacking an RI stage, do occasionally produce RI's. Yet, children acquiring Italian or Spanish produce considerably fewer RI's, and stop producing RI's at an earlier age, than (for example) children acquiring French or German. Differences across languages are thus statistical rather than absolute, and in evaluating the Russian data, a point of comparison is needed. We will use longitudinal production data for Italian and Polish as our points of comparison, drawing in part on Guasti's previous (1993) analyses of Italian. Polish is morphologically similar to Russian but differs in allowing null subjects of the Italian type. Thus, both Hyams & Hoekstra and Rhee & Wexler predict a relative absence of the RI stage in Polish, as well as Italian.

2. Objectives
The objectives of our study were the following:

(1) Longitudinal investigation of Root Infinitives (RI's) in the spontaneous speech of children acquiring Russian.

(2) Comparison of Russian results with Guasti's (1993) analyses of the longitudinal data for three children acquiring Italian, and our own analyses of the data from three children acquiring Polish and from three additional children acquiring Italian.

(3) Examination of Russian data for a possible interaction between finiteness of verb and overtness of subject (cf. Weverink 1989, Phillips 1996).

3. Method
Our evidence concerning the acquisition of Russian comes from longitudinal studies of two children. The transcripts of Varya, a Russian child studied by E. Protassova of the Russian Academy of Education, are available through the CHILDES database (MacWhinney & Snow 1985, 1990). The second child, Tanya, is a monolingual Russian child whose family is temporarily living in the United States. Tanya has been videotaped regularly in her home,

near the University of Connecticut, and the resulting tapes transcribed in the CHAT format for compatibility with the CHILDES CLAN software. The CLAN MLU program was used to obtain word-based MLUs. Varya's six transcripts range in age from 1;6 (MLU 2.6 words) to 2;4 (MLU 4.3 words). Tanya's fifteen transcripts range in age from 2;5 (MLU 1.7 words) to 3;2 (MLU 3.0 words).

Our principal comparison values for Italian are drawn directly from (Guasti 1993), and are based on the CHILDES corpora for Martina (ages 1;8-2;7, MLU 1.6-2.7 words), Diana (ages 1;8-2;6, MLU 2.5-6.0 words), and Guglielmo (ages 2;2-2;11, MLU 2.2-4.8 words). These corpora were contributed to the CHILDES database by the Calambrone group in Pisa (Cipriani et al., 1989). In addition to Guasti's data, we obtained new data by performing our own preliminary analysis of the transcripts for three more Italian children recently added to the Calambrone corpus: Viola (ages 1;11-2;10, MLU 1.7-2.7 words), Raffaello (ages 1;7-2;11, MLU 1.2-3.9 words), and Rosa (ages 1;7-3;3, MLU 1.3-3.2 words).

Comparison values for Polish are based on our own analyses of the CHILDES corpora for Marta, Bartosz, and Kubus. The Polish corpora were contributed to the CHILDES database by Richard Weist (Weist & Witkowska-Stadinik 1986). Thus far, we have analysed Transcripts 1-6 for Marta (ages 1;7-1;10, MLU 1.7-2.4 words), Transcripts 1-4 and 6 for Bartosz (ages 1;7-1;11, MLU 1.5-1.8 words), and Transcripts 1-7 for Kubus (ages 2;1-2;6,

MLU 2.1-2.8).

Each of the Russian and Polish children's utterances was hand-coded for finiteness of the main verb of the matrix clause. Additionally, each of the Russian children's utterances was hand-coded for overtness of the matrix subject. Our results for Russian, and Guasti's prior results for Italian, were plotted by age, by MLU of the corresponding transcript, and by "attained MLU." The *attained MLU* at a given age was the maximum of the MLUs for all the transcripts up to that age; thus, attained MLU, unlike simple MLU, never decreases over time.

4. Results

The Russian children produced far more RI's than did any of the Polish children. Out of 4,519 total utterances analysed for the three Polish children, at most nine utterances contained an RI. In contrast, in the 4,341 utterances for Varya, 103 of 1099 matrix verbs (9.4%) were RI's, and in the utterances for Tanya, 35 of 641 matrix verbs (5.5 %) were RI's. The children produced RI's in both "intensional" and "extensional" contexts, though intensional contexts were more common, at least in Varya's data (cf. Wijnen 1995).

Figures 1-3 illustrate the comparison of RI's in the two Russian children with those of the three Italian children (Martina, Guglielmo, and Diana) studied by Guasti. The graphs illustrate the rate of occurrence in three different ways: by age, by simple MLU, and
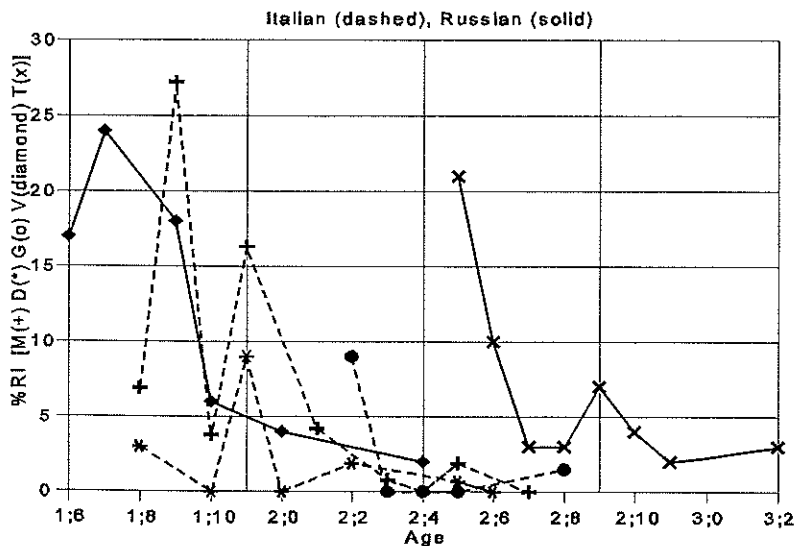

Figure 1: Root Infinitives by age.

by maximum attained MLU. Differences between Russian and Italian were not evident in the age-based comparison (Figure 1), but the MLU-based comparisons (Figures 2 and 3) revealed somewhat higher rates of RI's in the Russian children within the MLU range of 2.5 to 3.5 words. This difference was greater in the comparison based on attained MLU (Figure 3). Preliminary results for three additional children (Viola, Rosa, Raffaello) recently added to the Calambrone corpus indicate that the rate of RI production by these children during the MLU range of 2.5-3.5 words was no higher than for the three children previously examined by Guasti.

Finiteness of the verb interacted dramatically with overtness of the subject in Varya's data. Varya's use of overt subjects with finite verbs ranged from 31% (36:79) to 67% (149:75), but she used virtually no overt subjects with RI's (1:13, 0:37, 0:26, 1:13, 2:7, and 1:4 for the six transcripts studied). The finiteness/overtness contingency was significant (p<.05) for every transcript, by Chi-square test where applicable, or by modified sign test.

Varya 1 (Age 1;6, MLU 2.60 words)

	+FIN	-FIN
+SUBJ	30	1
-SUBJ	37	13

Varya 2 (Age 1;7, MLU 2.99 words)

	+FIN	-FIN
+SUBJ	56	0
-SUBJ	59	37

Varya 3 (Age 1;9, MLU 2.40 words)

	+FIN	-FIN
+SUBJ	36	0
-SUBJ	79	26

Varya 4 (Age 1;10, MLU 3.10 words)

	+FIN	-FIN
+SUBJ	131	1
-SUBJ	116	13

Varya 5 (Age 2;0, MLU 3.30 words)

	+FIN	-FIN
+SUBJ	143	2
-SUBJ	85	7

Varya 6 (Age 2;4, MLU 4.26 words)

	+FIN	-FIN
+SUBJ	149	1
-SUBJ	75	4

In contrast to Varya, Tanya showed little difference in her use of null/overt subjects with finite and non-finite (RI) verbs. Tanya's use of overt subjects with non-finite verbs was 34% overall, and 39% with finite verbs. (Ages are in the format, "years; months.days.")

Tanya 1 (Age 2;5.11, MLU 1.66 words)

	+FIN	-FIN
+SUBJ	4	0
-SUBJ	19	9

Tanya 2 (Age 2;5.24, MLU 1.59 words)

	+FIN	-FIN
+SUBJ	4	2
-SUBJ	14	0

Tanya 3 (Age 2;6.1, MLU 2.11 words)

	+FIN	-FIN
+SUBJ	17	2
-SUBJ	33	4

Tanya 4 (Age 2;6.8, MLU 2.35 words)

	+FIN	-FIN
+SUBJ	8	2
-SUBJ	14	0

Tanya 5 (Age 2;7.6, MLU 3.26 words)

	+FIN	-FIN
+SUBJ	15	0
-SUBJ	22	1

Tanya 6 (Age 2;7.26, MLU 2.58 words)

	+FIN	-FIN
+SUBJ	8	0
-SUBJ	16	1

Tanya 7 (Age 2;8.12, MLU 2.48 words)

	+FIN	-FIN
+SUBJ	24	0
-SUBJ	12	1

Tanya 8 (Age 2;8.25, MLU 2.35 words)

	+FIN	-FIN
+SUBJ	13	0
-SUBJ	18	0

Tanya 9 (Age 2;9.6, MLU 2.80 words)

	+FIN	-FIN
+SUBJ	21	1
-SUBJ	23	3

Tanya 10 (Age 2;10.11, MLU 2.28 words)

	+FIN	-FIN
+SUBJ	13	1
-SUBJ	30	1
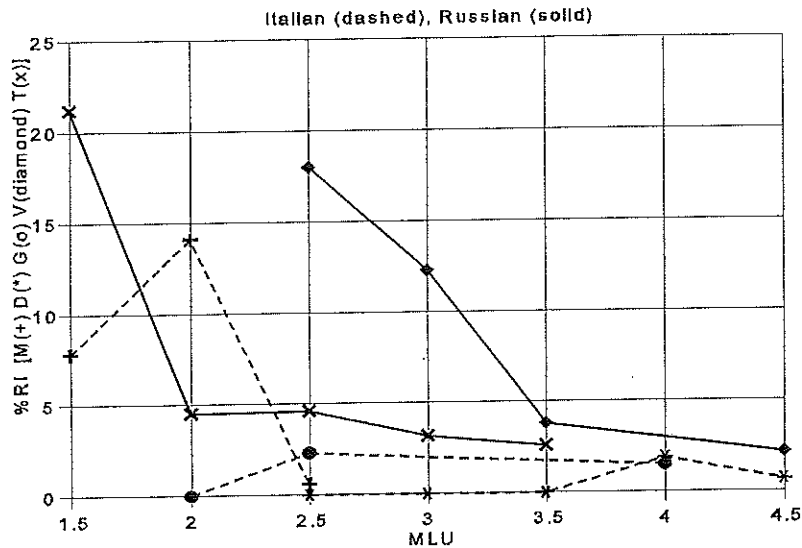
Figure 2: Root Infinitives by MLU

Tanya 11 (Age 2;10.27, MLU 2.25 words)

	+FIN	-FIN
+SUBJ	10	1
-SUBJ	55	0

Tanya 12 (Age 2;10.24, MLU 2.22 words)

	+FIN	-FIN
+SUBJ	9	0
-SUBJ	15	0

Tanya 13 (Age 2;11.1, MLU 2.50 words)

	+FIN	-FIN
+SUBJ	19	1
-SUBJ	31	1

Tanya 14 (Age 2;11.8, MLU 2.84 words)

	+FIN	-FIN
+SUBJ	28	1
-SUBJ	36	1

Tanya 15 (Age 3;1.23, MLU 2.99 words)

	+FIN	-FIN
+SUBJ	44	1
-SUBJ	30	1

5. Discussion

When transcripts are matched for simple MLU or for attained MLU, differences emerge between the Russian and Italian children. Within the MLU range of 2.5 to 3.5 words, Varya's RI production was substantially higher, and Tanya's RI production was somewhat higher, than that of any of the Italian children.

Among the Italian children, only one, Martina, produced a substantial number of RI's in her corpus. By the time Martina's MLU reached 2.5 words, however, her RI's accounted for only 0.6% of her matrix-clause main verbs. At the same MLU, Varya's rate of RI production was 18%, and Tanya's was 4.6%.

Differences between Russian and Polish were even more clear-cut. The Polish children produced virtually no RI's in the transcripts studied.

For Russian, an interaction between finiteness of the verb and overtness of the subject was also clear in Varya's data: The rate of overt subjects was far lower with RI's than with finite verbs. However, the same interaction was much less obvious, if present at all, in Tanya's data.

6. Conclusions

The principal implication of our findings is that more children need to be examined before drawing strong conclusions about cross-linguistic differences in the RI phenomenon. Possible differences emerge between the Russian and Italian children when they are matched by attained MLU, and to a lesser degree when matched by simple MLU, but these differences are completely obscured in an age-based
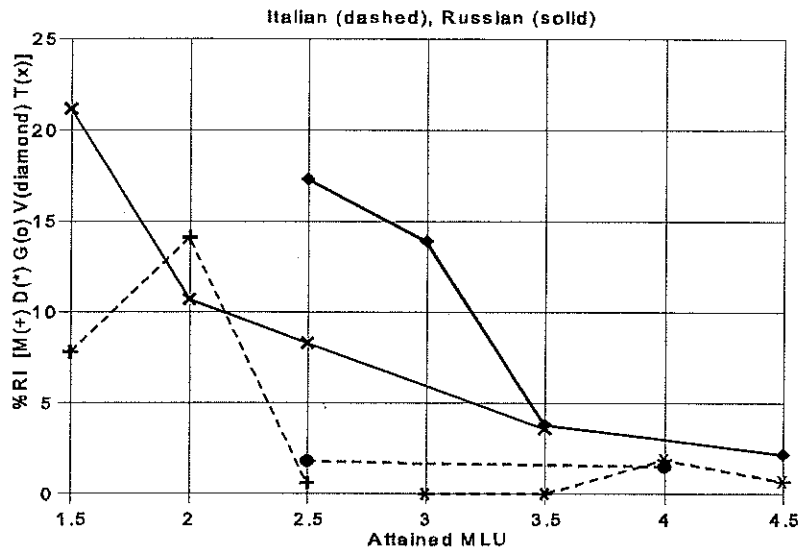
Italian (dashed), Russian (solid)

Figure 3: Root Infinitives by Attained MLU

comparison, and even in the MLU-based comparisons, the individual variability among children learning the same language is dramatic. Moreover, MLU is less than an ideal basis of comparison, given that transcribers often disagree on where an utterance ends, and given that MLU provides only a very general measure of linguistic development.

Differences between Polish and Russian are more robust than those between Italian and Russian, which suggests that comparison of closely related languages, forming near "minimal pairs," may be a useful research strategy. To the extent that Russian plausibly differs from both Polish and Italian in having a higher rate of RI's, our findings support Rhee & Wexler's hypothesis that the RI stage is found specifically in languages that lack the null subject option. On this interpretation of the findings, our results suggest, more specifically, that the unavailability of thematic null subjects is the relevant factor, because expletive subjects in Russian are in fact null.

Nonetheless, the possible difference in RI production between Russian and Italian needs to be interpreted with caution, because at best Tanya's production of RI's is only slightly higher than that of the Italian children, and then only within the MLU range of 2.5-3.5 words.

7. Acknowledgements

8. References

Chvany, C. (1973) Notes on "root" and "structure-preserving" in Russian. In *You Take the High Node and I'll Take the Low Node*, CLS Parasession 9, pp.252-290.

Cipriani, P., Pfanner, P., Chilosi, A., Cittadoni, L., Ciuti, A., Maccari, A., Pantano, N., Pfanner, L., Poli, P., Sarno, S., Bottari, P., Cappelli, G., Colombo, C., and Veneziano, E. (1989) *Protocolli diagnostici e terapeutici nello sviluppo e nella patologia del linguaggio*. (1/84 Italian Ministry of Health): Stella Maris Foundation.

Franks, S. (1995) *Parameters of Slavic Morphosyntax*. New York: Oxford University Press.

Guasti, M.-T. (1993) Verb syntax in Italian child grammar: Finite and non-finite verbs. *Language Acquisition* 3:1-40.

Hoekstra, T. and Hyams, N. (1995) Missing heads in child language. In C. Koster and F. Wijnen (eds.) *Proceedings of the Groningen Assembly on Language Development*.

MacWhinney, B. and Snow, C. (1985) The Child Language Data Exchange System. *Journal of Child Language* 12:271-296.

MacWhinney, B. and Snow, C. (1990) The Child

Language Data Exchange System: An update. *Journal of Child Language* 17:457-472.

Phillips, C. (1996) Root infinitives are finite. Some implications of cross-linguistic differences in two-year-olds' syntax. In A. Stringfellow et al. (eds.) *Proceedings of BUCLD 20*. Somerville, MA: Cascadilla Press.

Rhee, J. and Wexler, K. (1995) Optional infinitives in Hebrew. In C. Schuetze, J. Ganger, and K. Broihier (eds.) *Papers on Language Processing and Acquisition (MITWPL 26)*. Cambridge, MA: MIT Working Papers in Linguistics.

Weist, R. and Witkowska-Stadinik, K. (1986) Basic relations in child language and the word order myth. *International Journal of Psychology, 21,* 363-381.

Weverink, M. (1989) *The Subject in Relation to Inflection in Child Language.* M.A. Thesis, University of Utrecht, The Netherlands.

Wexler, K. (1994) Optional infinitives, head movement, and economy of derivation. In N. Hornstein and D. Lightfoot (eds.) *Verb Movement.* Cambridge, UK: Cambridge University Press.

Wijnen, F. (1995) Incremental acquisition of phrase structure. In J. Beckman (ed.) *Proceedings of NELS 25*. Amherst, MA: GLSA.