

**ACQUISITION OF NULL ANAPHORA IN BULGARIAN –  
A COPRUS-BASED STUDY<sup>1</sup>**

**1. Introduction**

The investigation of anaphoric relations is of pivotal relevance for the understanding of textual meaning. Anaphoric reference is on the one hand one of the main means of establishing textual coherence; on the other hand, anaphora are elusive linguistic devices as most of them, e.g. pronominal anaphora, are lexically underspecified and thus insufficient for reference determination.

From the perspective of theoretical linguistics, research on the properties of null anaphora – null pronouns and different types of ellipsis – has gained importance in the recent years, as null anaphors lack physical form but still have meaning<sup>2</sup>. Interest in the acquisition of null anaphora is driven by questions about the universal and language-specific properties of the syntax-semantics interface<sup>3</sup>.

In the present paper we argue that research which concentrates on the way children come to obey the syntactic well-formedness restrictions of subject drop is a necessary but not a sufficient basis for hypotheses concerning the acquisition process of null pronouns as

---

<sup>1</sup> The parallel spoken corpus of German, Russian and Bulgarian child narratives was developed within the project «Acquisition of intersentential anaphoric means» (ZAS, Berlin) sponsored by the German Science Foundation, GWZ 4/11-1. I'm grateful to the children and their teachers from the kindergardens of Vidin for their help in collecting the corpus data.

<sup>2</sup> See inter alia *Tomioka S.* The semantics of Japanese null pronouns and its cross-linguistic implications // *S. Winkler, K. Schwabe (eds.) The interfaces: Deriving and interpreting omitted structures.* Amsterdam, 2003. P. 321–340.

<sup>3</sup> *Lust B.C. (ed.) Studies in the acquisition of anaphora. Vol. I: Defining the constraints.* Dordrecht, 1986.

discourse-cohesive devices. We maintain that the coherence function of intersentential pronominal anaphora and particularly of null subject pronouns can be understood in greater depth only if the communicative demands in the production of connected discourse are taken into account. In this respect, annotated narrative corpora appear to be the most suitable source of statistically reliable and ecologically valid data. This claim is substantiated by contrasting experimental psycholinguistic findings with corpus-based evidence on the acquisition of null pronouns in Bulgarian.

## 2. Corpus-based studies in anaphora acquisition

Spoken longitudinal corpora have always been at the heart of child language<sup>1</sup>. The growing interest in the universal and language-specific principles and stages of the language acquisition process motivated the need for larger cross-linguistically comparable and statistically reliable data sources. Using the picture book story «Frog, where are you?», R. Berman, D. Slobin and their co-workers<sup>2</sup> have created a spoken narrative corpus suitable for cross-sectional and cross-linguistic studies in 5 typologically different languages<sup>3</sup> in order to gain insights into the acquisition course of temporal relations.

The acquisition of null pronouns has been studied extensively at the sentence level, mainly with a focus on the syntactic properties and licensing conditions of subject drop, but also from a processing point

---

<sup>1</sup> For an extensive overview of naturalistic corpora targeting discourse anaphora see: *Allen S., Skarabella B., Hughes M.* Using corpora to examine discourse effects in syntax // H. Behrens (ed.) *Corpora in language acquisition research: history, methods, perspectives*. Amsterdam, 2008. P. 99–138.

<sup>2</sup> *Berman R.A., Slobin D.I.* *Relating events in narrative: a cross-linguistic developmental study*. Hillsdale, NJ, 1994.

<sup>3</sup> The frog story data base has been extended to other languages like Russian by E. Protassova and to Korean. The frog-story corpus is available at: <http://childes.psy.cmu.edu/data/frogs>

of view<sup>1</sup>. In the following, the termini subject drop and null pronoun will be used synonymously in the syntactically motivated sense of J. Weissenborn (1992)<sup>2</sup>. Only utterances containing a finite verb which licenses subject drop by morphological agreement are taken to feature a null pronoun.

Previous research on the acquisition of pronominal anaphora has proven that narrative corpora are especially appropriate for looking at the correlation between the ability to produce a coherent story and to establish textual cohesion by means of anaphoric expressions. However, many of the studies focusing on pronominal anaphora concentrate on children at school age e.g. M. Hickmann (2003)<sup>3</sup>, E. Khoroujaia and L. Tolchinsky (2004)<sup>4</sup>.

The present paper provides evidence of the early steps in the acquisition of subject-drop in Bulgarian based on a corpus containing 300 videotaped narratives of pre-school children at age brackets of 3;0–3;6 – 4;0 – 4;6 – 5;0. Each age bracket includes 30 children who told 2 stories elicited in the picture-story design. The database used is part of an annotated parallel corpus of German, Russian and Bulgarian child narratives<sup>5</sup>.

---

<sup>1</sup> For discussion of this phenomenon in English based on longitudinal data and also on cross-linguistic and cross-sectional evidence see: *Valian V.* Syntactic subjects in the early speech of American and Italian children // *Cognition* 40, 1991. P. 21–81.

<sup>2</sup> *Weissenborn J.* Null subjects in early grammars: Implications for parameter setting theories // J. Weissenborn, H. Goodluck, T. Roeper (eds.): *Theoretical Issues in Language Acquisition*. Hillsdale, NJ, 1992. P. 269–299.

<sup>3</sup> *Hickmann M.* Children's discourse. Person, space and time across languages // *Cambridge Studies in Linguistics*, 98. Cambridge, 2003.

<sup>4</sup> *Khoroujaia E., Tolchinsky L.* Discursive constraints on the lexical realization in Spanish // R.A. Berman (ed.). *Language development across childhood and adolescence*. Amsterdam, 2004. P. 83–110.

<sup>5</sup> For a more detailed description of the corpus see: *Kuehnast M.* *Developing a corpus for contrastive studies of intersentential anaphora in*

The annotations include grammatical role and parts-of-speech tagging of anaphoric expressions like personal pronouns and demonstratives; indefinite and definite noun phrases. Null subjects are coded as such if licensed by verbal agreement morphology. The corpus is also tagged for referent features established in the literature on anaphora resolution as factors influencing the accessibility<sup>1</sup> of discourse entities. Referring expressions are annotated for animacy and grammatical role of the antecedents. Newness is marked in terms of morphological definiteness (definite articles and other determiners) and recency of mention (per number of propositions between antecedent and anaphor). Antecedents mentioned earlier than 5 propositions are considered as new. The corpus design provides data for cross-sectional studies allowing inferences about developmental stages in the use of overt and non-overt pronominal forms.

### 3. Null pronouns and topic marking in Bulgarian

Although there is a stable correlation between the accessibility status of discourse antecedents and the formal complexity of the anaphoric devices, developments in the field of natural language processing have shown that co-reference establishment is not a simple searching and eliminating procedure which easily finds the most appropriate referent. The formal properties of anaphoric expressions exhibit a substantial cross-linguistic variation and their role as information packaging devices depends on their functional distribution within a particular language.

Bulgarian is a strong pro-drop language in which the null pronoun contrasts with the personal pronouns. Besides obligatory subject

---

child language // Proceedings of the International Conference «Corpus linguistics 2006». St. Petersburg, 2006. P. 224–229.

<sup>1</sup> *Ariel M.* Accessibility theory: an overview // Sanders T., Schilperood J., Spooren W. (eds.). Text representation: linguistic and psycholinguistic aspects. Amsterdam, 2001. P. 29–87.

drop in 1<sup>st</sup> and 2<sup>nd</sup> person in neutral contexts, 3<sup>rd</sup> person null pronoun is a way of marking a highly salient discourse referent. In conversation, a null pronoun may be used by the interlocutors without explicit antecedent introduction if the referent is in their physical environment and in their joint focus of attention. In written texts, null pronouns locally mark the thematic referent<sup>1</sup> and are a means to maintain a specific referent as a discourse theme<sup>2</sup>.

As a member of the *Balkansprachbund*, Bulgarian has lost its nominal case marking system but has acquired some other morphological features which directly relate to marking of information structure, such as enclitic definite articles and a system of full (long) and clitic (short) pronominal forms. The use of a short pronoun together with a full pronoun or with a nominal phrase is a phenomenon known in the literature as 'pronominal/clitic doubling' and is subject to pragmatic restrictions such as topic marking.

From the perspective of a child acquiring Bulgarian as a native language, two quite contrary options are available for topic marking. Topics are introduced by specific indefinite or definite nominal phrases, regularly strengthened by clitic doubling when realized in object positions. Topic continuity is signaled through null anaphora licensed by verbal agreement or by short pronouns. The next section provides some experimental and corpus-based evidence on the use of these options in the early acquisition stages of Bulgarian.

---

<sup>1</sup> *Maslov J.S.* Grammatika bolgarskogo jazyka. Moskva, 1981.

<sup>2</sup> For an extended analysis of information packaging based on oral and written corpora of Bulgarian see: *Leafgren J.* Degrees of Explicitness. Information structure and the packaging of Bulgarian subjects and objects. Amsterdam, 2002.

#### 4. Early usage of subject drop – a corpus-based analysis

The present analysis is based on narrative data from 30 Bulgarian monolingual children at the age bracket 3;0 (range 2;10 – 3;03) and 20 adult Bulgarian speakers (mean age 34). The corpus data is contrasted with findings from a combined production and comprehension experiment<sup>1</sup> conducted with the same subjects.

The experiment was carried out in an act-out design and included an elicited imitation task for null, personal and demonstrative pronouns and a disambiguation task targeting the influence of the antecedents' features animacy and grammatical role on co-reference preferences of young children.

The production findings reveal that the 3-year-olds preferably produce null pronouns, substituting 3<sup>rd</sup> person personal and demonstrative prompts by null anaphora. The children observe the syntactic well-formedness conditions for the production of null pronouns. This claim gains support from the target-like production of sentences containing a copula, in which the use of a null pronoun induces the obligatory inversion of the enclitic copula.

The comprehension findings suggest that the extended production of null pronouns does not combine with an interpretative regularity. While adult Bulgarian speakers reliably resolve null pronouns to the subject in contexts with two animate antecedents, no such preference is found for the children. In order to verify the production-comprehension asymmetry of null pronouns found in the experiment, a comparison with a corpus-based analysis of contextual factors influencing the use of null anaphora appears necessary.

The corpus data reveals the following distribution of referring expressions in the narratives of the 3-year-old Bulgarian children.

---

<sup>1</sup> For a detailed description of the experimental design see: *Kuehnast M.* Production and comprehension of intersentential pronominal reference in early child Bulgarian – an experimental study. *ZAS Papers in Linguistics* 48, 2007. P. 171–202.

*Table 1.* Distribution of referring expressions

	<b>Children (3;0)</b>	<b>Adults</b>
<b>Null pronouns</b>	11%	7%
<b>Personal pronouns</b>	9%	14%
<b>Demonstrative pronouns</b>	7%	0%
<b>Indefinite NPs</b>	31%	7%
<b>Definite NPs</b>	42%	72%

The children tend to describe the content of each picture instead of telling a coherent story. Often they just name the referents without employing clausal structures. As a result, the children produce a significantly higher number of indefinite nominal phrases for the introduction of referents which appear to be new for them. Related to this pattern is the use of demonstrative pronouns for deictic pointing to referents in the pictures. Notwithstanding the high frequency of referent identification without action descriptions, around 15% of the children were found to maintain reference to a single protagonist, partly or fully abstracting from the presence of other referents. We interpret this descriptive strategy as a prerequisite for the ability to produce a coherent story organized around a main protagonist.

The tendency to establish one highly salient referent as a global discourse theme is mirrored in the way children make use of subject drop. While adults seem to observe locality restrictions (87% of the utterances with subject drop are adjacent to the antecedent-containing clause), children also use null pronouns to talk about antecedents mentioned 3 to 5 propositions earlier. In 9% of the cases, null pronouns referred to characters not explicitly introduced in the discourse. Note, however, that this is an option available in conversation if the referent is highly salient for the interlocutors. Children seem to over-generalize this usage in story telling. They employ a null anaphor to refer to a protagonist whose representation seems to be highly activated in their discourse model, but they ignore the felicity condition for the use of null subjects, namely that the focus of attention status must

be shared by the communication partners. In a similar way, the children use personal pronouns for co-reference establishment with referents not introduced in the previous discourse (13%).

With respect to the accessibility status of antecedents, we found that children and adults use subject drop to refer to animate antecedents in 99% of the cases. This result is related to the plots of the stories, which do not feature inanimate entities as active participants; rather, inanimate objects are employed in background descriptions. Animacy appears to be a basic prerequisite for the availability of the antecedent as a discourse topic from the very beginning of anaphora acquisition, and remains a main factor in adult anaphora resolution.

For adult Bulgarians, subject status is to be a reliable indicator of the probability that a subsequent co-reference relation will be realized by means of a null pronoun – 94% of the produced null anaphors are resolved to subject antecedents. No such correlation was found in the data of the 3 year olds – only 56% of the null pronouns also feature subject antecedents. The corpus analysis thus confirms the experimental results showing that at this developmental stage, the grammatical role of the antecedent is not yet established as a reliable disambiguation factor.

## **5. Conclusion**

The reliability and the ecological validity of corpus-based studies were illustrated against the background of specific questions concerning the early stages of null pronoun acquisition in Bulgarian as a strong subject-drop language. Age homogeneous narrative corpora combine the virtues of experimental and naturalistic child language data. Qualitative analyses based on such genre-consistent samples of connected speech allow for a deeper understanding of the mechanisms guiding the acquisition of the 3<sup>rd</sup> person null anaphor as a device which maintains an antecedent as a discourse theme.