To appear in *Journal of Child Language*

Running Head: HOW TO MEASURE DEVELOPMENT IN CORPORA?

How to measure development in corpora?

An association strength approach

Sabine Stoll
Max Planck Institute for
Evolutionary Anthropology, Leipzig

Stefan Th. Gries
University of California,
Santa Barbara

[*] This paper was mainly written during the second author's stays at the Max Planck Institute for Evolutionary Anthropology. We warmly thank Elena Lieven, Michael Tomasello and Bernard Comrie for providing such a stimulating and fascinating work environment. We thank Roland Meyer, who has done a wonderful job in developing a morphological tagger for Russian, and Tatjana Welikanowa and Madelaine Taoubi for help with coding. We are grateful to Balthasar Bickel, Elena Lieven and two anonymous reviewers for valuable comments.

ABSTRACT

In this paper we propose a method for characterizing development in large longitudinal corpora. The method has the following three features: (i) it suggests how to represent development without assuming predefined stages; (ii) it includes caretaker speech/child-directed speech; (iii) it uses statistical association measures for investigating co-occurrence data. We exemplify the implementation of these proposals with data on the acquisition of the patterning of tense and grammatical aspect of four Russian children. The method, however, is suitable for a wide range of other acquisition questions as well.

INTRODUCTION

The study of language development in corpora involves changes over time in the frequencies of occurrence of some linguistic variable or, alternatively, changes of co-occurrence of two or more variables. The way development is characterized, however, varies from researcher to researcher which renders comparisons across studies difficult. In this paper we propose a method to characterize development and exemplify this proposal on the basis of the development of tense/grammatical aspect correlations in a longitudinal corpus of Russian children.

Often in developmental research recordings are grouped into stages before analyzing the variable of interest. This is problematic for several reasons. First, any grouping of the data – i.e., the transformation of an otherwise rather continuous variable into a very small number of groups – results in the loss of information and preserves, at best, the ordinal information of the group (cf. Baayen, 2004: Section 2). Second, it is unclear on what basis to group the data. Age (as in the study on aspect by Li, Maher, Newark, & Hurley, 2001) is often not a reliable predictor of morphosyntactic development, which is basically why Brown (1973) and, after him, many others relied on mean length of utterance (MLU) as a more useful indicator of stages (cf. Bloom, Lifter & Hafitz, 1980; Shirai & Andersen, 1995 for studies on aspect). However, MLU values are notoriously variable and unstable. We show this in Fig. 1, in which the solid line represents MLU values (in words) of 66 recordings (lasting approx. 1 hour each) of Child 5 from the Stoll corpus of Russian language acquisition (Stoll, unpublished data). The child's age is expressed in decimal format on the left *y*-axis such that, for example, 2;6.0 is expressed as 2.5. The lower dashed line plots the sizes of the standard errors against the right *y*-axis at a higher resolution. Given the large variability exhibited in the data, it is probably fair to assume that unless an extremely explicit procedure is used, different researchers are unlikely to recognize the same MLU-based stages (cf. Gries & Stoll, to appear, for an algorithm solving this problem).

Figure 1

Finally, even if the classification based on MLU values were more reliable, it is not always obvious what is gained by such a classification (cf. Gries & Stoll, to appear), because, if, for instance, we are interested in tense/aspect marking, there is no obvious reason why the stages should be based on age, MLU or some other measure rather than on the phenomenon of interest i.e. tense/aspect markings (as in Aksu-Koç, 1998).

Another problem in the study of many phenomena and also in the study of tense/aspect is that we know very little about how the data of the children and the caretakers compare. For instance, there are several studies of aspect that analyze the data of the caretakers in addition to the data of the children but they all group the data into different kinds of stages so that the resolution is very coarse-grained and little direct comparison is made. Stephany's (1985) study of tense/aspect in Greek child language was the first to compare the correlations found in child-directed speech to those found in child speech. Her study is among the most comprehensive and she compares child-directed speech of the four mothers in her corpus to the output of the children, finding that the two are distributionally very similar. In the mothers' speech 96% of all past forms are perfective, compared to 100% of the children's past forms. However, few other studies perform similarly comprehensive comparisons. Shirai and Andersen (1995) investigate data of three English children but only compare the children and their caretakers at a single MLU stage, thus development cannot be traced. Aksu-Koç (1998) includes child-directed speech for one of her children and only a small interval is covered (child's age: 1;3.3-1;11.10). Li et al. (2001) provide fine-grained results by looking at how Vendlerian classes (lexical aspect) and tense-aspect marking develop over time, but they still group the data into year-long stages and thus lose precision.

A final issue concerns the way developmental trends are assessed. Simple percentages (as in Shirai & Andersen, 1995) and linear correlations are probably the most widely used measures. However, linear correlations such as product-moment correlations or linear regression make assumptions that are not always met (for instance, bivariate normal distribution and normally distributed residuals with similar variances). Li (2002) uses both simple chi-square tests and multifactorial log linear models to study the correlation between Aktionsart, aspect markers in Mandarin Chinese, and age. This method does justice to the multifactorial nature of the phenomenon but the age-based stages still give rise to the above-mentioned problems.

In the following, we make several methodological suggestions for studying development in language acquisition that address the shortcomings discussed above, focusing on quantitative analyses of distributional data from corpora:

(i.) to avoid a loss of information due to grouping, we make use of all individual data points;

(ii.) to better capture the covariation of the two variables, here tense and grammatical aspect, we use an association measure instead of cross-tabulation of pooled data (Cramer's *V*, cf. Bortz, Lienert & Boehnke, 1990: Section 8.1);

(iii.) to better compare how development takes place, we compare association statistics for children and caretakers; and

(iv.) to track developmental patterns in a noisy data set, we use smoothing and advanced regression techniques instead of simple linear regression models.

The proposed method is illustrated with a case study on the acquisition of Russian tense-aspect. However, it can be applied to a wide range of developmental questions that are based on the co-occurrence of two grammatical or lexical elements such as the use of nouns vs. pronouns in different constructions to test for preferred argument structure to give just one example.

CASE STUDY: TENSE-ASPECT CORRELATIONS IN RUSSIAN

Studies on a wide variety of languages have shown that in early acquisition there is a strong correlation between grammatical aspect (as a formal category, such as for instance perfective vs. imperfective) and tense, as well as between lexical aspect (*Aktionsarten*, such as for instance telics vs. duratives) and tense (for summaries see Li & Shirai, 2000; Weist, 2002). The correlation between tense and grammatical aspect as observed in the literature can be characterized as follows: ‘verbs with past tense marking 'for a bounded value' of grammatical aspect (e.g. perfective) are very likely to be telic whereas verbs with non-past marking for an unbounded value of grammatical aspect (e.g. imperfective or progressive) are likely to be non-telic’ (cf. Shirai, Slobin, & Weist, 1998: 246). Results concerning the strength of the association and its development over time vary, which may be due to the different age ranges studied and the different methods used. In this Note we suggest how the strength of associations can be measured by explicit mathematical procedures as exemplified in a large corpus of Russian child language.

Aspect in Russian

Concerning grammatical aspect, Russian distinguishes between perfective and imperfective aspect. Every verb is either perfective or imperfective with a number of biaspectual verbs (cf. Chertkova, 1996 about the increasing number of biaspectual verbs in Russian). Perfective is the marked member of the opposition and can be defined as expressing the action 'as a total event summed up with reference to a single specific juncture' (Forsyth, 1970: 8) e.g. *On napisal pis'mo* 'He wrote the/a letter (i.e. the writing was completed). Thus perfective aspect in Russian activates a boundary of the event described and one of the possibilities is that the boundary corresponds with the concept of completion. Imperfective aspect is unmarked and may or may not refer to the boundaries of the action expressed by the verb e.g. *On pisal pis'mo* 'He wrote a/the letter' or 'He was writing a/the letter'. Grammatical aspect systematically interacts with tense. Russian has three tenses: past, present and future. Imperfective aspect occurs in all three tenses but perfective aspect has only a past tense form and a future tense (which has the same ending as the imperfective present tense). In interaction with aspect the tenses get their specific meaning, i.e. whether an action was or will be completed (in interaction with the perfective aspect) or simply was, is or will be ongoing (in interaction with the imperfective aspect).

Since our main concern here is methodological, we do not discuss tense/aspect patterning in detail (for a proposal for qualitatively interpreting the development of tense/aspect through the tracking of the development of individual predicates, see Weist, Pawlak, & Carapella, 2004). We therefore concentrate on the association between tense and grammatical aspect (rather than between tense and lexical aspect) because grammatical aspect is morphologically coded and thus accessible in a Russian corpus (Stoll unpubl. data). We also disregarded all imperatives and imperfective future forms since they are not relevant for the present study. Imperatives have aspect but no tense and imperfective future is an important tense in Russian but in this paper we concentrate on the question whether the interrelation with present tense and imperfective and past tense and perfective also holds in Russian child language and child directed speech.

METHOD

Data

Our case study is based on the caretaker/child interactions of four Russian children (two sibling pairs) taken from a longitudinal corpus of Russian language acquisition (cf. Stoll, unpubl. corpus)[1]. All four children are monolingual Russian children living in St. Petersburg, Russia. Child 3 and Child 5 are the target children of the longitudinal recordings. Child 4 is the older brother of Child 3 and Child 6 is the older brother of Child 5. The mothers were students and lived either in a communal apartment (Child 3 and 4) or in an extended family setting (Child 5 and 6); this means that sometimes several other caretakers were present during the recordings. The children were recorded weekly, mainly with video at their homes in free interaction with their caretakers. Child 3 and Child 4 were recorded together as were Child 5 and Child 6. Child 4 and Child 6 did not take part in all the recordings and to the same extent because they were not the target children of the longitudinal study but merely part of the natural environment of the target children. The earlier recordings of Child 3 were not part of this study since they did not contain any verbs. Child 4 was 3;1 at the first recordings and his tense aspect behavior was analysed for the present study. Child 6, who was 11 years at the beginning of the study served as a control child for the present study. The recordings consisted of undirected interactions and were made with a wide-angle lense without an experimenter present.

Procedure

The data were transcribed by a Russian native speaker, double-checked by the mother of the children, and morphologically tagged by an automated stochastic tagger (Meyer, 2003). The tagger is 92-94% correct overall, however, for verbs the rate is nearly 100% because of the near absence of ambiguous forms. Nevertheless the tags for each verb were checked manually by a linguist. The only mistakes found were a handful of names that were erroneously tagged as verbs but no other mistakes occurred. Table 1 summarizes the number of utterances of the four children and their caretakers. 'Caretakers' here is a cover term for all adult native speakers providing input to the child during the recordings.

Table 1

From each recording, we retrieved all verb forms produced by the child and his/her caretakers and extracted the code for the speaker and the annotations of tense and aspect for the verb forms. (All retrieval operations as well as statistical computations and plots were performed with R for Windows 2.4; cf. R Development Core Team 2006). Crucially and as mentioned above in (i.) no grouping of the data into stages was performed.

Methods

Since we are interested in the probabilistic association between tense and aspect in each individual recording (cf. suggestion (ii.)), we use a probabilistic measure of association to quantify the association, namely Cramer's *V*, as our most central statistic. Cramer's *V* is computed from a χ^2-statistic for contingency tables as shown in (1) where the expression 'min[no of rows, no of columns] represents the smallest dimension of table.

(1) Cramer's V / φ = $\sqrt{\frac{\chi^2}{n \cdot (\min[\text{no of rows, no of columns}] - 1)}}$

For 2-2 tables of the kind that we will report on in this paper, the denominator simplifies to n and Cramer's V is therefore equivalent to Pearson's contingency coefficient *phi*.[2] Both range from 0 to 1. It is close to 0 when aspect and tense are not correlated, and the closer it is to 1 the stronger the correlation. Note that (1) also means that Cramer's V is independent of the sample size, which allows us to compare the associations of tense and aspect in differently sized recordings.

We then cross-tabulated all present tense forms and all past tense forms of each recording for the child and, separately, for the caretakers so that – cf. suggestion (iii.) – pairwise comparisons are possible. For example, for the recording of Child 3 (at age 2;7.28) and her caretakers, the observed frequencies (and expected frequencies in parenthesis) provided in Table 2 and Table 3 were obtained. The χ^2-values for these tables are 11.25 and 31.036 respectively (at $df = 1$ both of these are highly significant), and Cramer's V for each table is 0.5 and 0.402 respectively. Analogous computations were performed for all recordings.

Table 2

Table 3

Fourth, for each child and respective caretakers, we generated a scatterplot of Cramer's V

values such that the recording time is on the *x*-axis and Cramer's *V* is on the *y*-axis applying statistical methods to characterize the observed developmental trends.

RESULTS

Child 3

Child 3 is the youngest child in this study and was just starting to use verbs in the sessions chosen for this analysis (1;11.28). Thus, given the findings in the literature, we expected to find the strongest developmental trend in this child. Fig. 2 gives the results for Child 3 and also shows linear regression lines and their 95% confidence bands.

Figure 2

In both the data of the child and in the data of the adults we find large variation in the correlation of tense and grammatical aspect over time. However, the range of occurring values differs strongly in the child and the adults. The data of Child 3 exhibit considerable variation (*max* = 1; *min* = 0.016; *mean* = 0.452; *variation coefficient* = 0.44).[3] The data of the adults, however, exhibit less heterogeneity (*max* = 0.559; *min* = 0.163; *mean* = 0.364; *variation coefficient* = 0.24). The Cramer's *V* values for Child 3 exhibit a decreasing trend ($R^2 = 0.15$; $F(1, 78) = 13.74$; $p < 0.001$; *intercept* = 0.816; *slope* = -0.1129) whereas the Cramer's *V* values for her caretakers do not ($R^2 = 0.02$; $F(1, 78) = 1.2$; $p = 0.276$; intercept = 0.415; *slope* = -0.0159). This is what we would expect: First, as Cramer's *V* for the caretakers is approximately 0.4, the caretakers exhibit the preferred correlation between tense and grammatical aspect observed in previous studies, but do not exhibit any development. Second, the child starts out with a conservative correlation between tense and grammatical aspect and this correlation weakens later in development.

However, problems with the linear regressions become obvious quickly. First, when the validity of the linear regression is tested, it turns out there are clear *U*-shaped relations between the residuals and the fitted values for both the child and her caretakers. Second, the regression lines leave a lot of variability unaccounted for. Third, the valid range of prediction for the linear regression is small: for the child, the regression equation predicts that Cramer's *V* will be slightly smaller than 0 at approximately age 7;3, which its mathematical properties render impossible. Thus, in order to better summarize the development of the Cramer's *V* values over time, we replaced the linear regression line with a smoother resulting from a locally weighted linear regression (as implemented in R; cf. Cleveland, 1979).

Figure 3

As shown in Fig. 3, the linear regression line does not represent the kind of curvature obtained in the child's data very well whereas the smoother picks out two markedly different developmental phases. The child moves from a very strong correlation of tense and grammatical aspect at the beginning to more flexible behavior with the strength of the correlation decreasing over time until approximately age 3. Around age 3, the curve flattens considerably and then strongly resembles the curve of the caretakers. While the child is more conservative in her marking than her caretakers for nearly all recordings, around 3;0, her tense-aspect patterning approximates that of the adults very closely. By contrast, given the (expected) absence of development for the caretakers, the smoothing curve results in the same interpretation as the straight linear regression line does for Fig. 2.

The overall developmental trend is also reflected clearly in the left panel of Fig. 4, where we plot the differences between the Cramer's *V* values of the child and her caretakers as vertical lines against the age of the child (using moving averages across three recordings). Going from left to right, the differences between the association strengths of Child 3 and her caretakers, decrease (the vertical bars become shorter) as the child approximates the distributional patterns of the adult more and more closely (the bars tend to center around the caretaker mean of around 0.4).

Figure 4

Finally, let us note that there is a way in which an extension of regular linear regressions may be useful, namely regression with breakpoints (cf. Baayen, 2008: Section 6.4 and Crawley, 2002: Chapter 22 for details about this approach). We iteratively split up the data at every individual recording into an early part and a late part and then computed linear regressions in which the dependent variable was the vector of Cramer's *V* values of the child, and the independent variable was the interaction between the child's age and an indicator variable that marks each age as being part of the early or the late part. For each of these regressions, we stored the model deviance and then chose the model whose breakpoint was smallest and after which only increasing deviances were found. The two regressions following from this – one before the breakpoint, one after it – are shown in the right panel of Fig. 4.

Model comparison shows that the breakpoint at this location is highly warranted: If the amount of variance the linear model with the breakpoint explains (R^2 = 0.266) is compared to that of the linear model without the breakpoint from above, it emerges that the regression with a breakpoint can explain significantly more variance: $F(1, 77) = 11.991$; $p < 0.001$. In addition, the result nearly perfectly replicates the results of the smoother: from shortly before 2;0 until approximately 3;0, there is a strong downward trend (note the correlation coefficient, which is much smaller than the one obtained for all of the data). As of 3;0, on the other hand, there is no more development and the slopes of regression lines of the both the child and her caretakers do not differ significantly from 0 anymore. The lack of a significant difference between the slopes of the regression lines after age 3 reflects the lack of development found for both child and caretakers. It does not, however, mean that the heights of the Cramer's *V* values do not differ significantly, which could be tested, e.g., with a paired Wilcoxon test. Both the smoother and the regression with breakpoints analysis reveal a bifurcation of the developmental data that simple linear summary statistics and premature groupings of the data might well have missed.

Child 4

Child 4's recordings begin at a later age (3;1.8), more specifically at an age at which Child 3 has already begun to approximate the caretakers' patterning, which is why we expect a less pronounced developmental trend. Consider Fig. 5, where again the left and right panels show the data for the child and his caretakers respectively.

Figure 5

On the one hand, again, we find considerable variation. However, the range of occurring values differs not only strongly in the child and the adults but also between Child 4 and Child 3.

The data of Child 4 exhibit considerable variation (*max* = 0.806; *min* = 0.037; *mean* = 0.419; *variation coefficient* = 0.31), but much less so than Child 3 as would be expected from the different age ranges. The data of the adults are again less heterogeneous (*max* = 0.645; *min* = 0.163; *mean* = 0.392; *variation coefficient* = 0.24).

On the other hand, as to the developmental pattern, our expectation is again confirmed: (i) we find a much less strong developmental slope than for Child 3, but (ii) as of age 3, the slopes of the smoothing curves for both Child 3 and Child 4 are largely identical, descending only slightly, and (iii) both the overall mean and the overall slope for Child 4 are virtually identical to those of the caretakers.

Child 5

Child 5 (2;3.17) is slightly older than Child 3, but younger than Child 4, and a relatively early talker (especially in terms of lexical development). It is, thus, difficult to formulate precise predictions. Consider Fig. 6.

Figure 6

These data differ from those for the first two children: Child 5 exhibits a relatively small degree of variation of the Cramer's *V* values (*max* = 0.719; *min* = 0.291; *mean* = 0.388; *variation coefficient* = 0.2). A further difference is that his caretakers' patterning is more heterogeneous (*max* = 0.638; *min* = 0.048; *mean* = 0.353; *variation coefficient* = 0.33). On the whole, the developmental pattern of Child 5 is less pronounced than that of Child 3 and more similar to that of Child 4. Still across virtually all recordings the child is again much more conservative than his caretakers, who, in turn, exhibit no developmental trend at all.

Child 6

As a control, we looked at the tense/aspect correlation of the 11-year-old brother of Child 5 (11;7.18). If our method is on the one hand sensitive enough to identify developmental patterns, but on the other hand not over-sensitive (such that one always obtains strong developmental curves and/or huge differences in terms of variation), then the data for Child 6 should resemble those of his caretakers very closely. This is exactly what we find, as is shown in Fig. 7.

Figure 7

Both the data of the child and the data of the adults are rather similar in terms of their overall variation across the sessions (Child 6: *max* = 0.794; *min* = 0.109; *mean* = 0.409; *variation coefficient* = 0.4, caretakers: *max* = 0.638; *min* = 0.048; *mean* = 0.337; *variation coefficient* = 0.37). Second, both curves do not exhibit any clear developmental pattern and are close to the results obtained for all previous adults just as would be expected for an 11-year-old child.

DISCUSSION

The focus of this Note was on the method we propose, however there are also some findings on the development of tense/aspect in Russian. As in previous studies, we found the expected overall correlation between tense and grammatical aspect for both children and adults. Thus, for both adults and children there was a strong correlation between tense and aspect. The correlation in the children's earlier recordings, however, was much stronger than in the adults'. Our study confirms that the strong correlation children exhibit in early development weakens over time to a degree that is close or identical to that of the adults. However, this study is the first to show this with an association strength method that indicates the exact correlations between tense and aspect. Since we also avoid grouping, the methodology yields a continuous evaluation of the acquisition process and this type of analysis further enables us to determine, with a high degree of precision, the point at which the child's data resembles that of their caretakers.

Some of our results differ from those of previous studies. Li et al. (2001: 130) state that 'there seem to be 'developmental' trends even in the parental input, in that the associations become weaker as the input age increases.' This does not correspond to our results; so we revisited their data. The correlation between tense and aspect in our data is compatible with the correlation between tense/aspect and Aktionsart in Li et al.'s data. However, in our data, there is no developmental trend in the adult data: the overall average of the Cramer's *V* values of all caretakers across all recordings is about 0.4 with a small degree of variation and no systematic change. In Li et al.'s data there is a slight developmental trend: one cannot make such comparisons on the basis of quoted chi-square values, but we computed effect sizes for their data and found Cramer's *V*s of 0.355, 0.27, and 0.242 for their input stages.

We can only speculate where the difference between their results and ours comes from, but there are two likely explanations. First, Li et al. studied the correlation between tense (past, present progressive, third person singular *s*) and *Aktionsart* (process, state, etc.) and these data may pattern slightly differently from our data on tense (past vs. non-past) and grammatical aspect (imperfective vs. perfective) in Russian. Second, unlike Li et al. we did not lump the data of the caretakers into yearlong stages. Not only is it unclear whether this division is motivated by something having to do with tense/aspect, this lumping may also hide significant differences among parts of the data that have been pooled. For example, Li et al. (2001: 130) state that '[e]xamining each input stage separately, we found the same strong associations between lexical aspect and grammatical morphology as [in the pooled data].' But this does not hold: the *strength* of the association obtained in their pooled data is much stronger since Cramer's *V* for the pooled data is 0.563 (rather than the above 0.355, 0.27, or 0.242). Thus, the sampling and successive grouping of the data has resulted in data that are quantitatively different and whose association strength is less homogeneous than assumed. We suggest that this underscores our reasoning against pooling and in favor of reporting association statistics for samples separately.

CONCLUSION

We discussed four methodological proposals to help characterize development in large longitudinal corpora. First, these proposals avoid a need for stages and thus an overall loss of information as well as the data loss and risks coming with grouping rather continuous data into few groups. The approach is therefore fine-grained, but on the other hand we have shown it is not overly powerful since it is able to generate the expected null results for the control child. Second, the patterning of the child-directed speech is included and serves as a reference against which the children's data can be compared. Third, the proposed improvements over simple (linear) summary statistics – smoothing methods as well as regressions with breakpoints – show that sometimes only more refined methods can reveal the strongest and most interesting patterns such as different developmental phases. Finally, instead of just reporting overall tables or percentages, we used a measure of association strength that allows to directly assess the issue at hand: the association between and co-occurrence of grammatical categories.

It is worth pointing out that the first three proposals can be applied regardless of whether the dependent variable is concerned with associated co-occurrence frequencies as in the present example or any other kind of quantitative dependent variable (such as frequencies, or percentages). Finally, the proposed method is not restricted to the pairing of tense and aspect. In fact, any aspect of grammatical development that involves the co-occurrence of two lexical or grammatical elements is eligible for analysis with the proposed association strength approach.

REFERENCES

Aksu-Koç, A. (1998). The role of input vs. universal predispositions in the emergence of tense-aspect morphology: evidence from Turkish. *First Language*, **18**, 255-80.

Baayen, R. H. (2004). Statistics in psycholinguistics: a critique of some current gold standards. *Mental Lexicon Working Papers,* **1**, 1-45.

Baayen, R.H. (to appear). *Analyzing linguistic data: a practical introduction to statistics*. Cambridge: Cambridge University Press.

Bloom, L., Lifter, K. & Hafitz, J. (1980). Semantics of verbs and the development of verb inflection in child language. *Language,* **56**, 386-412.

Bortz, J., Lienert, G. A. & Boehnke, K. (1990). *Verteilungsfreie Methoden in der Biostatistik*. Berlin: Springer.

Brown, R. (1973). *A first language: the early stages.* Cambridge, MA: Harvard University Press.

Chertkova, M. I. (1996). *Grammaticheskaia kategoriia vida v sovremennom russkom iazyke.* Moscow.

Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association,* **74**, 829-36.

Crawley, M. J. (2002). *Statistical computing: an introduction to data analysis using S-Plus*. Chichester: John Wiley and Sons.

Forsyth, J. (1970). *A grammar of aspect: usage and meaning of the Russian verb.* Cambridge: Cambridge University Press.

Gries, St. Th., & Stoll, S. (to appear). Finding developmental groups in acquisition data: variability-based neighbor clustering. *Journal of Quantitative Linguistics.*

Li, P. (2002). Loglinear models for the analysis of language acquisition data. *Cognitive Science,* **3**, 27-41.

Li, P. & Shirai, Y. (2000). *The acquisition of lexical and grammatical aspect*. Berlin/New York: Mouton de Gruyter.

Li, P., Maher, S., Newmark, E. & Hurley, J. (2001). The role of parental input in the acquisition of tense-aspect morphology. *Journal of Cognitive Science,* **2**, 119-43.

Meyer, R. (2003). Halbautomatische morphosyntaktische Annotation russischer Texte. In R. Hammel,. & L. Geist (eds.). *Linguistische Beiträge zur Slavistik aus Deutschland und Österreich. X. JungslavistInnen-Treffen, Berlin 2001*. München: Sagner, 92-105.

R Development Core Team. (2006). *R: a language and environment for statistical computing*. R Foundation for Statistical. Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Shirai, Y., & Andersen, R. W. (1995). The acquisition of tense-aspect morphology: A prototype account. *Language,* **71**, 743-62.

Shirai, Y., Slobin, D. I. & Weist, R. M. (1998). Introduction. In Y. Shirai, D. I. Slobin & R. M. Weist (eds.) *The acquisition of tense-aspect morphology (Special Issue of First Language, vol. 18)*. Buckinghamshire: Alpha Academic.

Stephany, U. (1985). *Aspekt, Tempus und Modalität. Zur Entwicklung der Verbalgrammatik in der Neugriechischen Kindersprache*. Tübingen: Narr.

Stoll, S. (unpubl. data). A corpus of Russian first language acquisition.

Weist, R. M. (2002). The first language acquisition of tense and aspect. In R. Salaberry & Y. Shirai, (eds.), *The L2 acquisition of tense-aspect morphology.* Amsterdam, Philadelphia: John Benjamin.

Weist, R. M., Pawlak, A. & Carapella, J. (2004). Syntactic-semantic interface in the acquisition of verb morphology, *Journal of Child Language,* **31**, 31-60.

Table 1: Summary of the analyzed recordings

Child	Age span	Recordings	Child utterances	Caretaker utterances
Child 3	1;11.28 - 4;3.12	80	6,796	31,687
Child 4	3;1.8 - 6;7.12	117	19,652	50,611
Child 5	2;3.17 - 5;6.26	66	11,447	20,749
Child 6	11;7.18 - 13;11.1	42	5,524	12,697

Table 2: Tense × aspect correlation of Child 3 at age 2;7.28

Child 3	Non-past	Past	Totals
Imperfective	25	5	30
Perfective	5	10	15
Totals	30	15	45

Table 3: Tense × aspect correlation of the caretakers of Child 3 at age 2;7.28

Caretakers Child 3	Non-past	Past	Totals
Imperfective	112	14	126
Perfective	35	31	66
Totals	147	45	192

Fig. 1: MLUs of Child 3 between 1;11.28 and 4;03.12

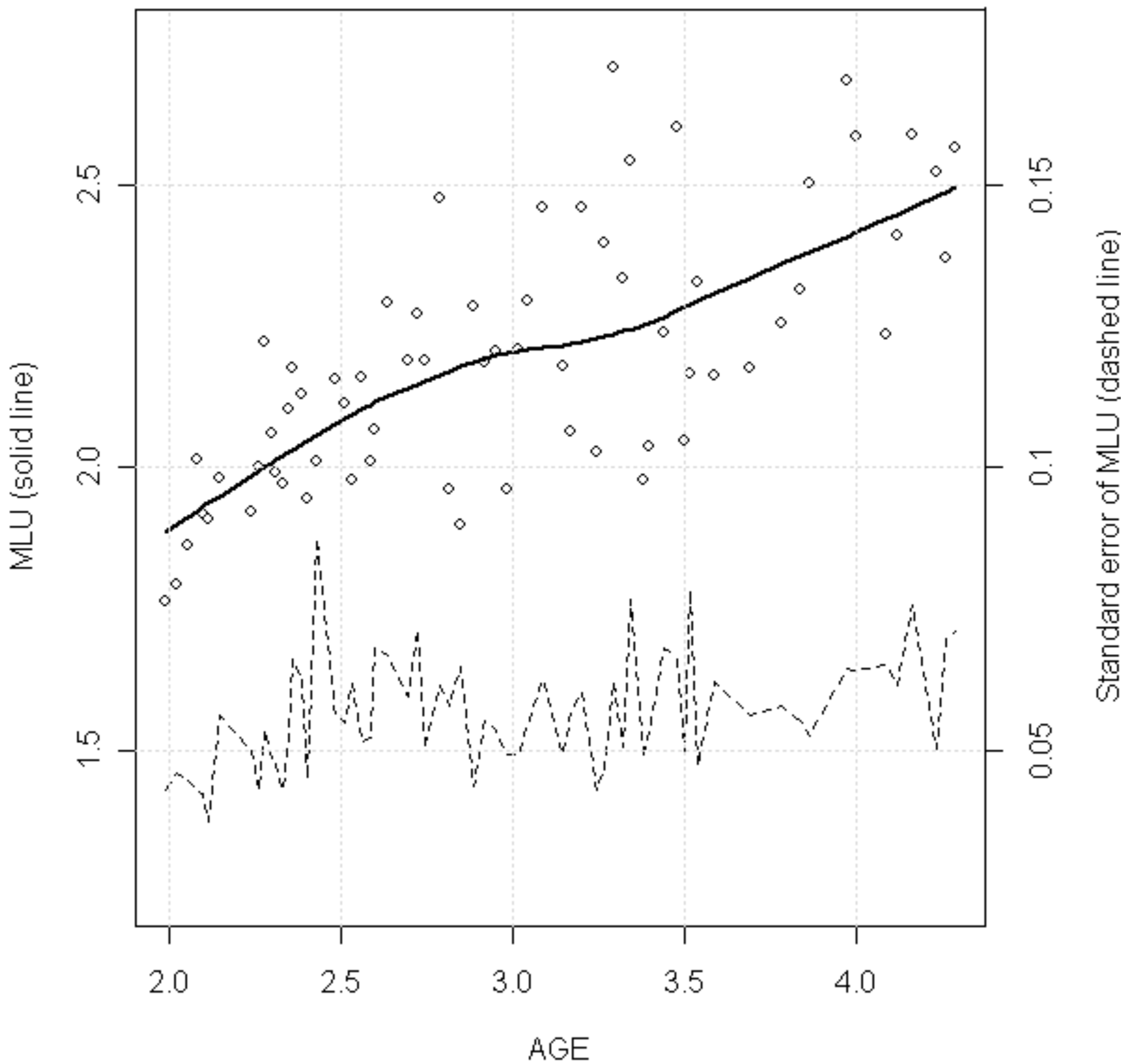


Fig. 2: The tense × grammatical aspect correlation of Child 3 (left) and her caretakers (right): linear regressions and confidence intervals

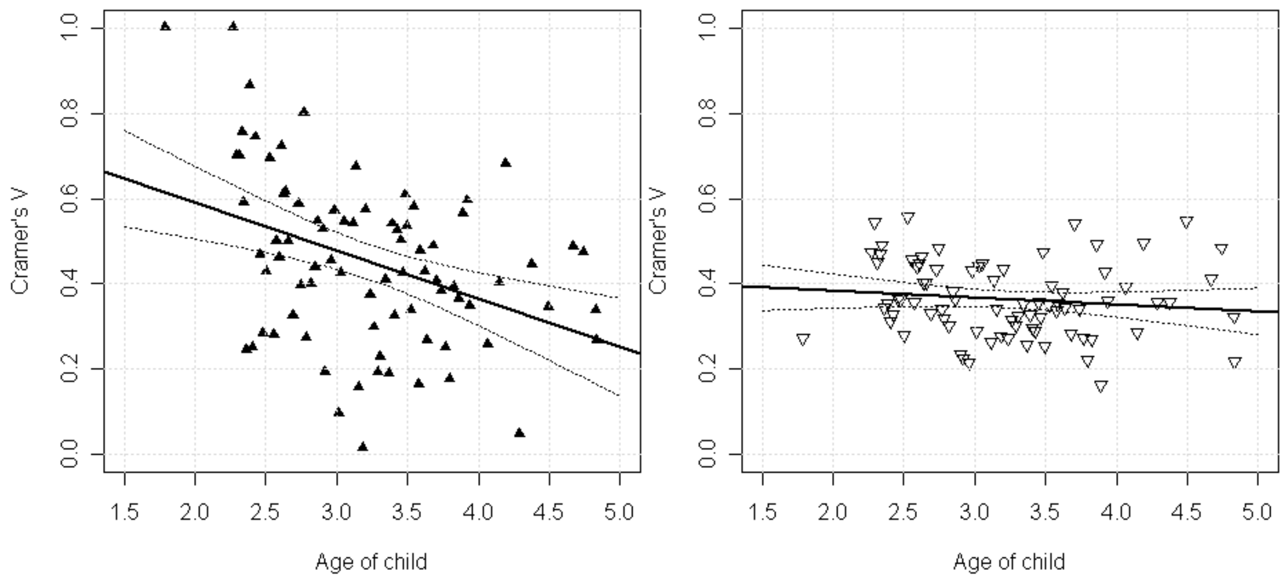


Fig. 3: The tense × grammatical aspect correlation of Child 3 (left) and her caretakers (right): confidence intervals of linear regression and locally weighted regression line

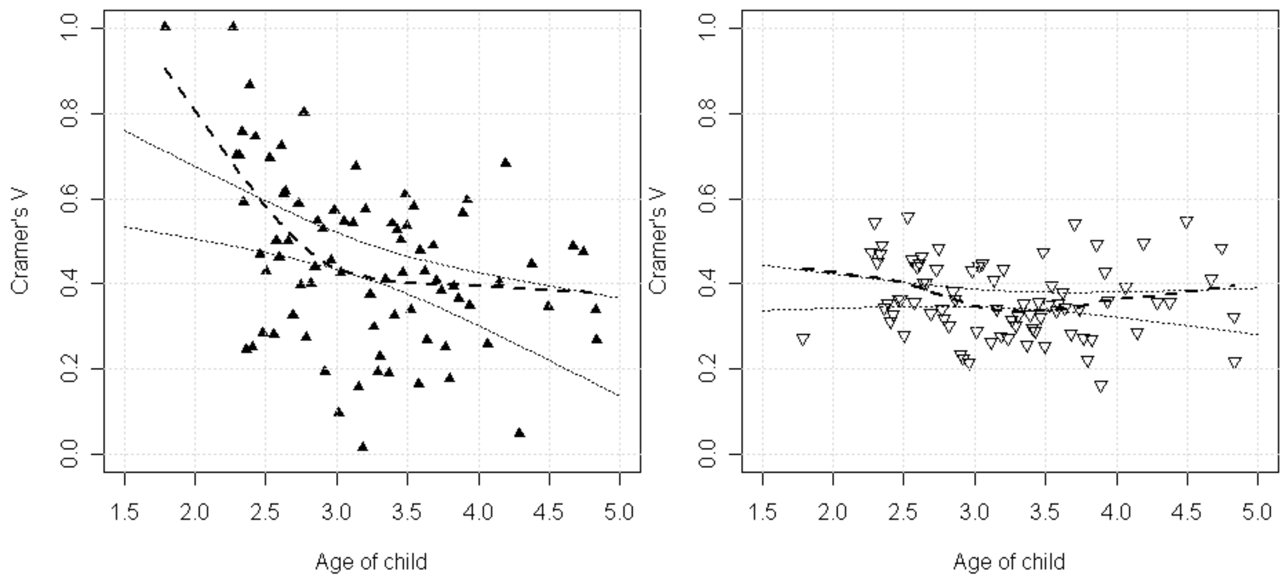


Fig. 4: The tense × grammatical aspect correlation of Child 3 (left) and her caretakers (right): differences between Cramer's V values and regression with breakpoints

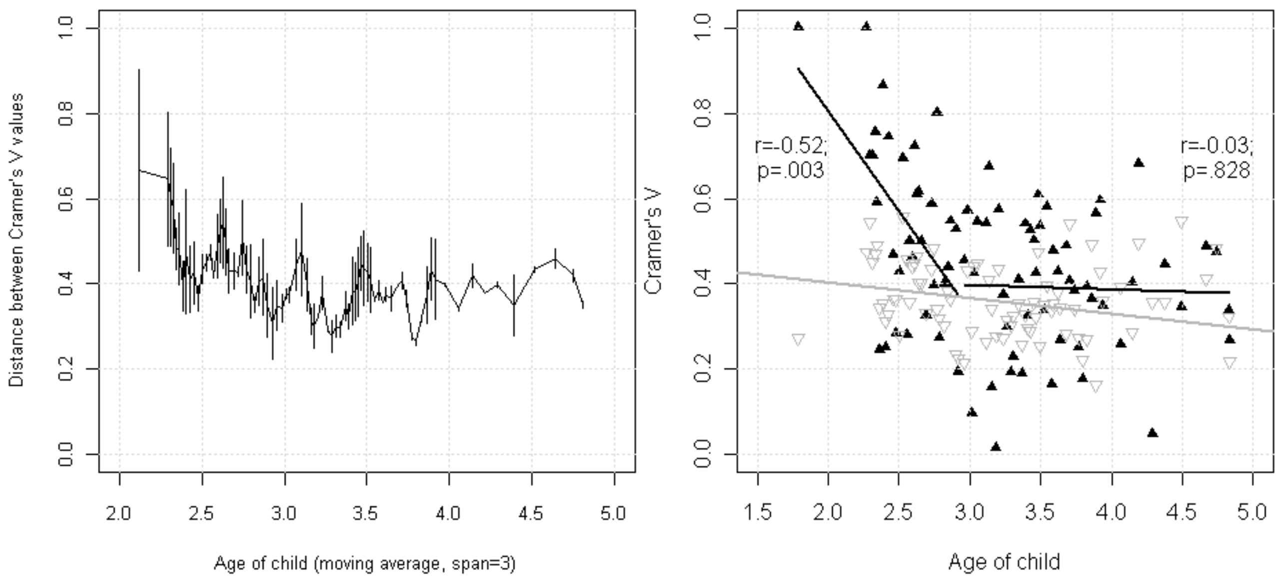


Fig. 5: The tense × grammatical aspect correlation of Child 4 (left) and his caretakers (right): confidence intervals of linear regression and locally weighted regression line

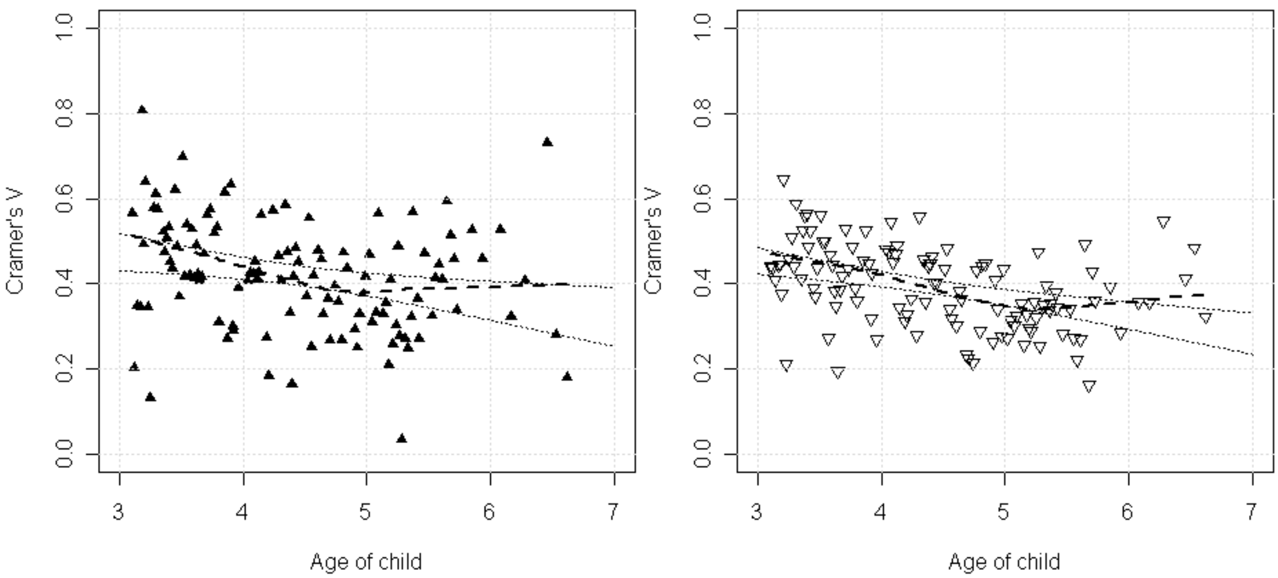


Fig. 6: The tense × grammatical aspect correlation of Child 5 (left) and his caretakers (right): confidence intervals of linear regression and locally weighted regression line

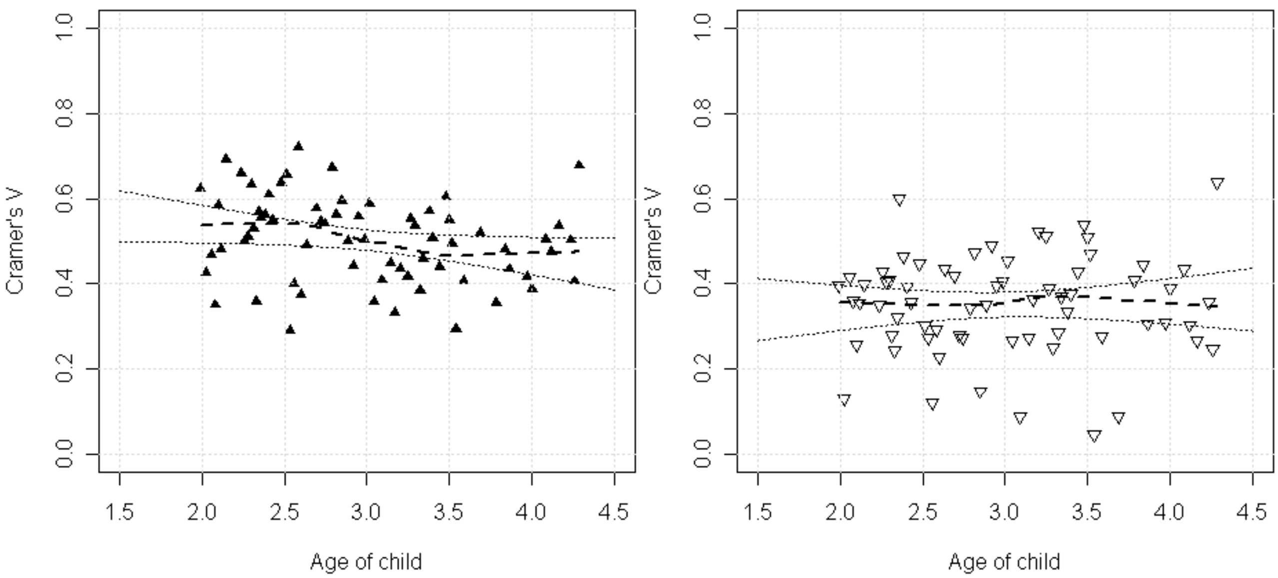
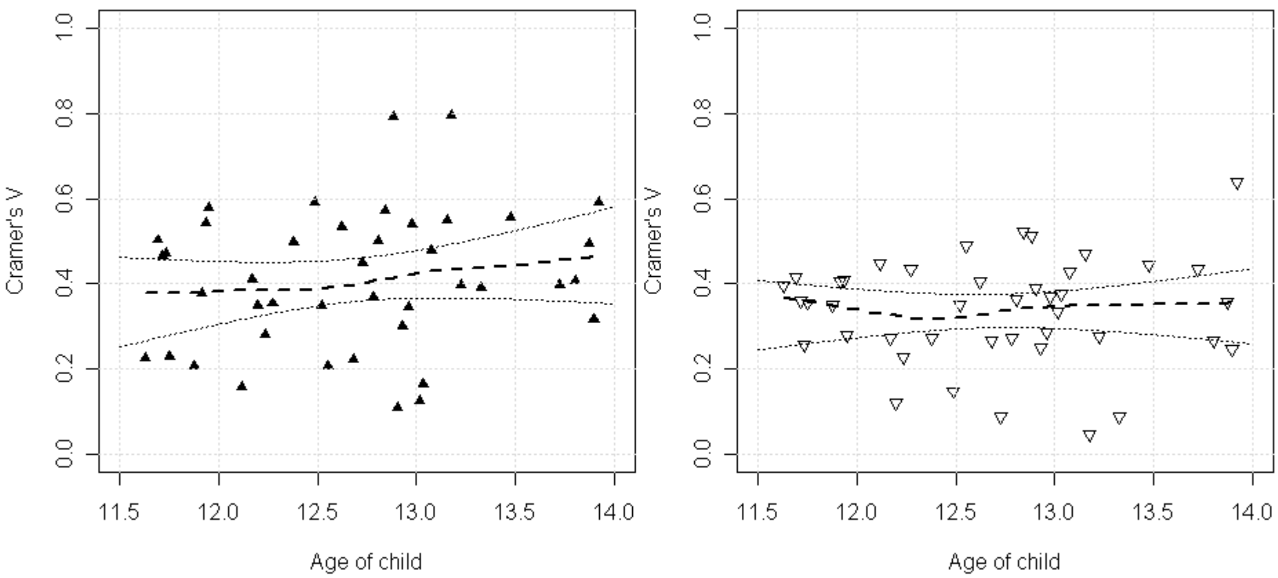


Fig. 7: The tense × grammatical aspect correlation of Child 6 (left) and his caretakers (right): confidence intervals of linear regression and locally weighted regression line



Footnotes:

[1] In the longitudinal study six children were studied. Child 1 and Child 2 of the longitudinal study are not part of the present study.

[2] This dependency of Cramer's *V* entails that *V*'s interpretation at least with regard to statistical significance is subject to the same constraints as all chi-square tests. Thus, one could not use the measure to test a 2 x 3 tense aspect-system having an empty cell.

[3] The variation coefficient is a measure of dispersion that is better suited for comparing different dispersions across measures and samples than the standard deviation, as it is a function of the absolute amount of dispersion, rather than being normalized to one as in the standard deviation. It is computed by dividing the standard deviation of a distribution by its mean.