

## Computer-Based Assessment of Academic Language Skills – the “CALSy”-Tool

**Jens Siemon, Joana Duarte, Ingrid Gogolin**

LiMA Research Cluster, University of Hamburg (Germany)  
[jsiemon@ibw.uni-hamburg.de](mailto:jsiemon@ibw.uni-hamburg.de); [joana.duarte@uni-hamburg.de](mailto:joana.duarte@uni-hamburg.de)

### Abstract

*Academic language as defined by Cummins (2000 [1]) and Schleppegrell (2004 [2]) is a central linguistic register for reaching higher school outcomes in educational systems, particularly in the case of pupils with a migration background – Gogolin and Lange (2011 [3]), Leseman, Scheele et al. (2009 [4]), Schleppegrell (2004). Furthermore, research on school quality in multi-ethnic constellations has shown that an explicit focus on academic language skills and language learning correlates with other indicators for higher school quality - Gogolin (2009 [5]), Helmke and Klieme (2008 [6]), Helmke, Goebel et al. (2003 [7]), Prenzel and Allolio-Näcke (2006 [8]).*

*Although there is a growing number of language assessment instruments that can be used in multilingual teaching settings (i.e. available in different language versions, for e.g.) and account for some of the phenomena of multilingual speakers (Döll and Dirim 2010 [9]), very few allow an assessment of the academic language skills essential to obtain higher school outcomes. The federal program FörMig – Support for Pupils with a Migration Background – developed between 2004 and 2009 three language assessment instruments for the languages German, Turkish and Russian with a focus on academic skills (Gantefort and Roth 2010 [10]). However, the increasing demand of educational institutions, as well as of research on language development, makes it reasonable to integrate computer-based automated analysis of language into assessment systems.*

*The paper reports on the CALSy<sup>1</sup>-Tool, based on a rule driven, automated tagging and parsing system. It analyses written or transcribed text in given content areas. As result, the tool generates statistical data on several aspects of vocabulary, morphology and syntactical structure, including special aspects of academic language usage. These results can be used for scientific interpretation, but also as a source for assessment of language development. The aim is to provide educational professionals with a tool able to determine German academic language proficiency of multilingual speakers with a minimum effort.*

### 1. Introduction: language, migration and school outcomes

Large monitoring studies have repeatedly highlighted the disparities between the performance of pupils with a migration background and that of their monolingual peers across Europe. This performance gap is more accentuated in Germany than in most other OECD countries (Schnepf 2007 [11]). Although it has been continuously decreasing, the PISA 2009 study (Klieme et al. 2010 [12]) still shows significant differences in the performance of first and second generation pupils with a migration background and monolingual German pupils.

Several approaches from different disciplines have tried to account for this performance gap (Diefenbach 2010 [13]). Most explanations, however, identify the lack of age appropriate proficiency in the language of instruction as one of the central aspects leading to educational disparities. Since the

---

<sup>1</sup> Computer-Based Assessment of Academic Language.

first PISA results, a large amount of programmes and projects has started with the aim of promoting language support of pupils with a migration background. However, language promotion programmes and concepts focused on general linguistic abilities have not produced the expected results. The federal programme FörMig<sup>2</sup> suggested a change of focus to the particular language register called "Bildungssprache" (similar to the concept of cognitive academic language proficiency, see Cummins 2000), which, when explicitly conveyed to pupils with migration background, seems to produce visible learning effects (FÖRMIG/Gogolin 2010 [14]). "Bildungssprache" is, on the one hand, used and expected by educational institutions, occurring in classroom talk and teacher discourse in general, but also in text books, and on the other, it is expected from "successful" pupils who succeed attain higher educational outcomes. Recent research (Leseman et al. 2007 [15]; 2009) confirmed that general linguistic proficiency is not as much responsible for higher school outcomes as the particular register of the academic language proficiency. Furthermore, research on school quality in multi-ethnic constellations has shown that an explicit focus on academic language skills and language learning correlates with other indicators for higher school quality - Gogolin (2009), Helmke, Goebel et al. (2003), Prenzel and Allolio-Näcke (2006).

From the educational linguistics' perspective, as well as from the point of view of language assessment, it is therefore central that proficiency in academic language be thoroughly identified, so that language support programmes for multilingual pupils can be more accurately planned and conducted. From a sociocultural theory perspective (Vygotsky 1986 [16]), the so-called "zone of proximal development", defined as the gap between what learners already know and what they can achieve when provided with adequate educational support, can only be reached when the current state of knowledge is accurately determined. The aim of the CALSy-Tool is thus to determine academic language proficiency of multilingual speakers in educational settings, in order to (a) support language learning planning and (b) monitor language development across time through repeated application.

## 2. Assessing multilingual speakers: challenges for the CALSy-Tool

The development (and attrition) of multilingual language proficiencies depends on language environment and language use; the multilingual language environments are themselves subject to change (Schrauf 2009 [17]). Multilingual speech of children and youngsters is often characterized by phenomena such as code-switching and the use of emerging linguistic forms, which therefore are not conform to normative assumptions of the target language. However, most of these emerging forms or code-switching phenomena are signs of high language knowledge and are mostly happen either through transfer from the other language or through generalization of rules<sup>3</sup>.

Another issue related to assessing multilingual speakers derives from testing procedures and reliability of results. Most data on multilingual speakers is collected via self-assessment scales and tests, thus not being based on productive language data. Results from a recent study, however, found that correlations between active speech data and self-assessment scales were in many cases not significant or rather moderate (Rauch, Jurecka and Hesse, 2010 [18]). In fact, other studies have highlighted different biases in self-assessment of second language speakers<sup>4</sup>.

Based on this controversy, the CALSy-Tool is based on productive oral and written language samples as means of assessing academic language proficiency.

---

<sup>2</sup> Support for Pupils with a Migration Background

<sup>3</sup> An example would be to use the simple past form "he flyed", instead of "he flew". This form suggests that the speaker is aware of the regular ending for the simple past and applied it to the irregular verb "fly". Such forms are also common in monolingual language acquisition.

<sup>4</sup> MacIntyre, Noels and Clément, 1997 studied the role of language anxiety; Masgoret and Gardner, 2003 attitudes and motivation, Collentine and Freed, 2004, identity issues, for example.

### 3. The language assessment tool CALSy

The CALSy-Tool was developed based on the method of the profile analysis (Clahsen, 1985 [19]). Profile analysis methods belong to the category of diagnostic analytical tools and thus differ from so-called standardized tests. They allow assessment of second language development at greater depth and in a more holistic manner than standardized tests focused on very specific language competencies are able to do. Data collection is based on picture-based elicitation through the use of visual prompts of– for example a comic strip for a narrative elicitation or a series of images representing a scientific experiment for elicitation of description and interpretative language acts. Participants must either accomplish the task orally (when they are too young to write) or in written form. Oral data are recorded and transcribed before they are analysed, while written data are typed out and then analysed.

The CALSy-Tool analyses language samples of multilingual pupils in three different age groups; the 5-6 year-olds, 10-11 year-olds and the 14-15 year-olds. It was tested in cooperation with the federal project FörMig which developed these three diagnostic instruments for the language pairs, German-Turkish and German-Russian (Gantefort and Roth 2010). The instrument for the 5-6 year-olds – HAVAS 5 – is a picture-based elicitation whereby children must narrate a small story. The samples are recorded, transcribed and then fed into the tool. The aim is to identify emerging academic language skills and determine proficiency in German in the areas of vocabulary, morphology and syntax. The instrument for the 10-11 year-olds equally elicits a narrative text in a written form and assesses academic language skills as well as proficiency in the same areas as HAVAS. The instrument for the 14-15 year-olds elicits a more technical text of written instruction for the construction of a boomerang and assesses, in addition to the categories of the previous instruments, technical vocabulary. The written samples of the last two instruments are transcribed according to specific CALSy rules and then fed into the tool.

Categories for analyses include six main areas of language knowledge which have been tested, updated and validated within the FörMig project (Klinger et al. 2008 [20]). The choice of categories was based on the concept of language proficiency for the purpose of language diagnostics in German as suggested by Ehlich (2005 [21]).

The CALSy software (Siemon 2001 [22]; 2003 [23]) is able to carry out a part of the assessment work for the German version of the instruments. It is therefore a semi-automated software for language diagnostic. From the categories mentioned above, CALSy is able to process morphology, syntax, vocabulary, verbal form and position and academic language skills with a 95% hit rate (Döll, Roth and Siemon 2009 [24]). Task accomplishment, which is not language category but rather of cognitive nature must be analysed separately.

CALSy is a parsing system for text and syntax analysis based on head-driven phrase structure grammar (HPSG – Pollard and Sag 1994 [25]). The analysis of the texts of natural languages by means of EDP systems has been a topic of research since decades and substantial progress was achieved, particularly for the English language. Basically for the whole analysis process of such a system are three operations: tokenizing, tagging and parsing. Task of the tokenizers is to sequence an input chain (text). For further processing, word and sentence borders must be recognized as well as abbreviations identified. Current control and statistically based tokenizers already work very reliably. Error rates, for instance, concerning sentence borders recognition lay under one percent (Klatt and Bohnet 2004 [26]).

After tokenizing, the tagging process begins. Here meta-information for each token (e.g. word class) is identified and added. Klatt (2005 [27]) introduced a control-based multistage PEP-Tagger whose mistake rate with 0.11 percent excels the correctness rate of known statistical systems.

In the third operation, the parsing, the syntactic construction of a given sentence and its morphological structure are determined. For this purpose, the Parser compares the construction of the input to the grammar deposited in the program. In relation to the chosen approach and efficiency existing systems differ substantially. Klatt (2005) points out to the fact that, nevertheless, the system descriptions documented in the literature are often only superficial and that attempts to show the current state of research in this area are impaired by the vagueness of the situation. Moreover, Ruland (2001 [28]) draws attention to the problems of the differently used evaluation masses which work as a comparison of the systems' efficiency.

Results of analysis can be presented in two forms: (a) as table (for e.g. Excel or SPSS) which can be used for further analysis and be added to a corpus of existing samples; (b) as form where a brief summary of the pupil is presented, enabling more detailed language planning and monitoring.

#### 4. Conclusions

At the moment no automated language assessment tool exists in Germany for the group of the multilingual speakers and taking into account the particular language phenomena produced by this group. Furthermore, no tool is specifically directed for the use in educational settings and as basis for the planning of adequate language learning. In addition, automated assessment of academic language skills is a fairly innovative approach, which arose from the found correlation between proficiency in this register and the attainment of higher school outcomes (for a summary of the discussion see Duarte 2011 [29]).

As educational institutions in westernized urban areas growingly receive more multilingual pupils, the need for automated tools which determine language skills will also rise. Furthermore, schools and kindergartens also complain about the lack of specifically trained personnel to analyse language samples and perform planning (Bosselmann 2005 [30]). Once it is fully tested, CALSy can thus be used in these contexts. In addition, versions in Russian and Turkish are being planned.

#### References

- [1] Cummins, J. (2000) Language, power and pedagogy. Bilingual children in the crossfire. Clevedon: Multilingual Matters.
- [2] Schleppegrell, M.J. (2004) The language of Schooling. A Functional Linguistics Perspective. Mahwah, New Jersey: Lawrence Erlbaum Associates Publishers.
- [3] Gogolin, I. and Lange, I. (2011) Bildungssprache und durchgängige Sprachbildung. In: Fürstenau, S. und Gomolla, M. Migration und schulischer Wandel: Mehrsprachigkeit. Wiesbaden: VS Verlag, Pp. 107-127.
- [4] Leseman, P., Scheele, A et al. (2009) Bilingual development in early childhood and the languages used at home: competition for scarce resources? In: Gogolin, I and Neumann, U. Streitfall Zweisprachigkeit – The bilingualism controversy. Wiesbaden: VS Verlag, Pp. 289-316.
- [5] Gogolin, I. (2009): "Bildungssprache" - The Importance of Teaching Language in Every School Subject. In: Tajmel, Tanja/ Starl, Klaus (eds.): Science Education Unlimited. Approaches to Equal Opportunities in Learning Science. Münster u.a. (Waxmann), pp 91 - 102.
- [6] Helmke, A./Klieme, E. (2008): Unterricht und Entwicklung sprachlicher Kompetenzen. In: DESI-Konsortium (Hrsg.): Unterricht und Kompetenzerwerb in Deutsch und Englisch. Ergebnisse der DESI-Studie. Weinheim/Basel: Beltz. Pp. 301-312.

- [7] Helmke, A., Göbel, K., Hosenfeld, I., Schrader, F.W., Vo, T. & Wagner, W. (2003). Zur Rolle des Unterrichts im Projekt DESI. *Empirische Pädagogik*, 17 (3). Pp.396-411.
- [8] Prenzel, Manfred & Allolio-Näcke, Lars (Hrsg.) (2006). *Untersuchungen zur Bildungsqualität von Schule. Abschlussbericht des DFG-Schwerpunktprogramms BiQua*. Münster: Waxmann.
- [9] Döll, M. und Dirim, I. (2011) Mehrsprachigkeit in der Sprachdiagnostik. In: Sara Fürstenau und Mechthild Gomolla (Hrsg.): *Migration und Schulischer Wandel: Mehrsprachigkeit*. Wiesbaden: VS Verlag. Pp. 153-168.
- [10] Gantefort, C. & Roth, H.-J. (2010). Sprachdiagnostische Grundlagen für die Förderung bildungssprachlicher Fähigkeiten. In: *Zeitschrift für Erziehungswissenschaft*, Heft 13. Wiesbaden: VS Verlag für Sozialwissenschaften, S. 573-591.
- [11] Schnepf, S.V. (2007), 'Immigrants' Educational Disadvantage: An Examination Across Ten Countries and Three Surveys', *Journal of Population Economics*. Pp. 527-545.
- [12] Klieme, e. et al. (2019) PISA 2009. Bilanz nach einem Jahrzehnt. Münster: Waxmann Verlag.
- [13] Diefenbach, H. (2010): *Kinder und Jugendliche aus Migrantenfamilien im deutschen Bildungssystem. Erklärungen und empirische Befunde*. Wiesbaden: VS Verlag.
- [14] FÖRMIG / Gogolin, I. (2010) Abschlussbericht des BLK-Programms FÖRMIG. Mimeo, University of Hamburg.
- [15] Leseman, P. et al. (2007) Home Literacy as a Special Language Environment to Prepare Children for School. In: *Zeitschrift für Erziehungswissenschaft*, Heft 3/2007. Pp. 334-355.
- [16] Vygotsky, L. (1986). *Thought and language*. Cambridge, MA: The MIT Press.
- [17] Schrauf, Robert W. (2009). Longitudinal designs in studies of multilingualism. In DeBot, K. and Schrauf, R.W. (eds). *Language development over the lifespan (245-270)*. Routledge. Jersey: Lawrence Erlbaum Associates, Inc.
- [18] Göbel, Kerstin; Vieluf, Svenja; Hesse, Hermann-Günter: Die Sprachentransferunterstützung im Deutsch- und Englischunterricht bei Schülerinnen und Schülern unterschiedlicher Sprachlernerfahrung In: *Zeitschrift für Pädagogik. Beih.*, (2010) 55, 101-122
- [19] Clahsen, Harald (1985) Profiling second language development: A procedure for assessing L2 proficiency. In: Hyltenstam, K. & Pienemann, M. (eds.) *Modelling and Assessing Second Language Acquisition*. Clevedon: Multilingual Matters, 283-331
- [20] Klinger, T., K. Schwippert und B. Leiblein (Hg.), 2008: *Evaluation im Modellprogramm FörMig. Planung und Realisierung eines Evaluationskonzepts*. Münster: Waxmann.
- [21] Konrad Ehlich (2005) Anforderungen an Verfahren der regelmäßigen Sprachstandsfeststellung als Grundlage für die frühe und individuelle Sprachförderung von Kindern mit und ohne Migrationshintergrund. Eine Expertise für das Bundesministerium für Bildung und Forschung. Bonn/Berlin: BMBF
- [22] Siemon, J. (2001). Eine multimediale Betriebserkundung und die inhaltsanalytische Messung resultierender Lerneffekte. In H. Heid, G. Minnameier & E. Wuttke (Hrsg.), *Fortschritte in der Berufsbildung?* (S. 179-185). Stuttgart: Steiner.
- [23] Siemon, J. (2003). *Evaluation eines komplexen Lehr-Lern-Arrangements - Eine netzwerk- und inhaltsanalytische Studie am Beispiel der Einführung in ein Modellunternehmen*. Wiesbaden: DUV.
- [24] Döll, M., Roth, H.-J. and Siemon, J. (2009) Computergestützte Analyse der gesprochenen Kindersprache? Entwicklung und Erprobung eines parsergestützten Sprachtools für das Hamburger Verfahren zur Analyse des Sprachstands Fünfjähriger (HAVAS 5). In: Drorit Lengyel, Hans H. Reich, Hans-Joachim Roth und Marion Döll (Hrsg.): *Von der Sprachdiagnose zur Sprachförderung. FörMig Edition Band 5*. Münster: Waxmann.
- [25] Pollard, C. & Sag, I. A. (1994). *Head-driven phrase structure grammar*. Stanford: Center for the study of Language and Information.

- [26] Klatt, S. & Bohnet, B. (2004). You don't have to think twice if you carefully tokenize. In: 1st International Joint Conference on Natural Language Processing (IJCNLP-04), Hainan (China)
- [27] Klatt, S. (2005). Kombinierbare Textanalyseverfahren für die Korpusannotation und Informationsextraktion. Aachen (Shaker)
- [28] Ruland, T.N. (2004). Ein hybrider Ansatz zur Strukturanalyse gesprochener Spontansprache. Sankt Augustin: infix.
- [29] Duarte, J. (2011) Bilingual language proficiency. A comparative Study. Münster: Waxmann.
- [30] Bosselmann, U. (2005): HAVAS: Hamburger Sprachstandserhebung am Schulanfang. In: Bartnitzky, H. / Speck-Hamdan, A. (Hrsg.): Deutsch als Zweitsprache lernen. Beiträge zur Reform der Grundschule. Band 120. Frankfurt (Grundschulverband), 132-139.