

Processability Approaches to Language Acquisition Research & Teaching

5

Developing, Modelling and Assessing Second Languages

EDITED BY
Jörg-U. Keßler,
Anke Lenzing
and Mathias Liebner

John Benjamins Publishing Company

Developing, Modelling and Assessing Second Languages

Processability Approaches to Language Acquisition Research & Teaching (PALART)

ISSN 2210-6480

Processability Theory (PT) as developed by Manfred Pienemann is a prominent theory of second language acquisition. PT serves as a framework for a wide range of research covering issues such as L2 processing, interlanguage variation, typological effects on SLA, L1 transfer, linguistic profiling and L2 assessment, stabilisation/fossilisation and teachability. The PALART series serves as a platform for making current research within the PT framework and its application to measurement and teaching, as well as the interdisciplinary discussion of PT accessible to both researchers and graduate students in the field. PALART is designed to provide a thematic platform for the presentation of current high-quality work within the PT framework. The thematic scope of the series reflects the wide scope of theoretical, empirical and practical aspects of PT.

For an overview of all books published in this series, please see
<http://benjamins.com/catalog/palart>

Editors

Manfred Pienemann
University of Paderborn &
Newcastle University

Bruno Di Biase
Western Sydney University

Jörg-U. Keßler
Ludwigsburg University of
Education

Advisory Board

Camilla Bettoni
Verona University, Italy

Jonas Granfeldt
Lund University, Sweden

Angela Hahn
Ludwig-Maximilian University, Munich,
Germany

Gisela Håkansson
Lund University, Sweden

Satomi Kawaguchi
Western Sydney University, Australia

Anke Lenzing
Paderborn University, Germany

Ingo Plag
Heinrich-Heine-Universität Düsseldorf,
Germany

Matti Rahkonen
Jyväskylä University, Finland

Yanyin Zhang
Australian National University, Australia

Volume 5

Developing, Modelling and Assessing Second Languages
Edited by Jörg-U. Keßler, Anke Lenzing and Mathias Liebner

Developing, Modelling and Assessing Second Languages

Edited by

Jörg-U. Keßler

Ludwigsburg University of Education

Anke Lenzing

Paderborn University

Mathias Liebner

Carl-von-Ossietzky Gymnasium Bonn

John Benjamins Publishing Company

Amsterdam / Philadelphia



The paper used in this publication meets the minimum requirements of the American National Standard for Information Sciences – Permanence of Paper for Printed Library Materials, ANSI Z39.48-1984.

DOI 10.1075/palart.5

**Cataloging-in-Publication Data available from Library of Congress:
LCCN 2016004379 (PRINT) / 2016012812 (E-BOOK)**

ISBN 978 90 272 0305 2 (HB)

ISBN 978 90 272 6719 1 (E-BOOK)

© 2016 – John Benjamins B.V.

No part of this book may be reproduced in any form, by print, photoprint, microfilm, or any other means, without written permission from the publisher.

John Benjamins Publishing Company · <https://benjamins.com>

Table of contents

Acknowledgments	VII
Modelling and assessing second language acquisition: 30 years onwards <i>Anke Lenzing</i>	IX
About this book	XV
Part I: Theory Development	
The development of argument structure in the initial L2 mental grammatical system <i>Anke Lenzing</i>	3
Question constructions, argument mapping, and vocabulary development in English L2 by Japanese speakers: A cross-sectional study <i>Satomi Kawaguchi</i>	35
Processability Theory and language development in children with Specific Language Impairment <i>Gisela Håkansson</i>	65
Testing the Developmentally Moderated Transfer Hypothesis: The initial state and the role of the L2 in L3 acquisition <i>Manfred Pienemann, Anke Lenzing & Jörg-U. Kefßler</i>	79
Part II: Theory Assessment	
The 'tense' issue: Variable past tense marking by advanced end-state Chinese speakers of L2 English <i>Yanyin Zhang & Bo Liu</i>	101
Acquisition as a gradual process: Second language development in the EFL classroom <i>Jana Roos</i>	121
Psychometric approaches to language testing and linguistic profiling – A complementary relationship? <i>Katharina Hagenfeld</i>	135

Assessing linguistic levels of L2 English in primary school programs <i>Esther Maier, Lea Neubauer, Katharina Ponto, Stefanie Couve de Murville & Kristin Kersten</i>	163
Diagnosing L2-English in the communicative EFL Classroom: A task-based approach to individual and developmentally moderated focus on form in a meaning-focused setting <i>Jörg-U. Keßler & Mathias Liebner</i>	193
The cognitive processes elicited by L2 listening test tasks – A validation study <i>Henning Rossa</i>	207
About the Authors	239
Index	243

Acknowledgements

We would like to thank the following people for their support in writing and editing this book:

Kristof Baten (Ghent), Bruno Di Biase (Sydney), Angela Hahn (Munich), Gisela Håkansson (Lund and Halden), Magdalena Keltsch (Ludwigsburg), Alexander Kuuskowski (Ludwigsburg), Patricia Leplae (Amsterdam), Howard Nicholas (Melbourne), Manfred Pienemann (Paderborn), Jana Roos (Paderborn), Kees Vaes (Amsterdam), three anonymous reviewers and all our (applied) linguistics students in Paderborn and Ludwigsburg.

Additionally, we would like to thank the Forschungsförderungsstelle and the English Department (both Ludwigsburg University of Education) for their generous financial support for proofreading.

Jörg-U. Keßler, Anke Lenzing and Mathias Liebner
Ludwigsburg/ Paderborn/ Bonn, February 2016

Modelling and assessing second language acquisition

30 years onwards

Anke Lenzing

Paderborn University

The present volume brings together the work of a number of researchers working in the framework of PT and addresses several current issues in both theory development and theory application. The two-fold aim of the volume – the engagement with both theoretical and applied aspects of SLA – reflects a 30 year-old tradition of viewing SLA from a learner-centred perspective and relating insights into the L2 acquisition process, particularly those focussing on L2 developmental trajectories, to questions of language teaching and assessment.

As early as 1985, Kenneth Hyltenstam and Manfred Pienemann published an edited volume titled *Modelling and Assessing Second Language Acquisition*. Noticing that “knowledge gained [from SLA research] has not yet influenced the language teaching profession or the language classroom very much” (Hyltenstam & Pienemann 1985: 3), the overall aim of the volume was to bridge the gap between theory and practice and to explore the implications of SLA research for language teaching and assessment. The topic was approached from various angles by the different authors involved, focussing, for instance, on task-based language teaching (Long), different contexts of (school-based) SLA (Clyne), learner variation (Nicholas, Hyltenstam) and various aspects of assessing proficiency (Ingram, Hulstijn, Lapkin, Stölting, Clahsen, Fried). A core characteristic of the overall discussion was its learner-centred perspective. In line with this, a number of articles dealt with relationships between learning and teaching (e.g. Pienemann, Lightbown, Nicholas, etc.), and a major focus was on developmental sequences in SLA.

The title of the current volume – *Developing, modelling and assessing second languages* – makes reference to the 1985 publication. Co-ordinating approaches to addressing both theoretical and applied aspects of L2 acquisition is an essential part of bridging the gap between theory and practice and contributing to both teachers’ expectations concerning the L2 learning process as well as to improvements in assessment. Naturally, since the publication of the volume by Hyltenstam

and Pienemann, the field of SLA has grown extensively, not only in terms of the number of studies, but also with regard to the diversity of issues addressed. Today, SLA is a vibrant interdisciplinary research area, focussing on an expanding number of topics. Ortega (2015:245) identifies the following central areas addressed by SLA researchers: “(a) The nature of second language knowledge and language cognition, (b) the nature of interlanguage development, and the contributions of (c) knowledge of the first language (L1), (d) the linguistic environment, and (e) instruction.” These issues are approached from many and varied perspectives, ranging from cognitive and psycholinguistic to usage-based and sociocultural approaches and involving dramatic and diverse advances in theory development (see e.g. VanPatten & Williams 2015; Atkinson 2011).

However, despite the diversity of perspectives present in current SLA research and the resulting controversial debates on learning processes, potential explanations for specific phenomena and related research methodologies, the two issues of developmental sequences or trajectories and the application of research findings to language teaching and assessment continue to be considered to be highly relevant in SLA research.

The phenomenon of developmental sequences in SLA has a long research history, and the question of the existence of universal sequences in SLA is currently regarded as being “one of the central issues in understanding phenomena of second language acquisition” (Hulstijn 2015:1). Although discussed heatedly,¹ developmental sequences are regarded as an established finding in current SLA textbooks (see e.g. VanPatten & Williams 2015; Ortega 2009). Pienemann (2015:123) points out that “there has been a continuous focus on second language development in second language acquisition research for over 40 years and that there is clear empirical evidence for generalizable developmental patterns.” This focus includes, but is not limited to research within the area of PT (see Lenzing 2015; Pienemann 2015), which has explored L2 developmental trajectories in both theoretical and empirical terms since its articulation in 1998.

Since then, the field of research within the PT framework has also expanded considerably (see e.g. Pienemann & Keßler 2011; Pienemann & Lenzing 2015). The widened scope of the theory includes engagement with theoretical issues as well as with empirical research findings in a number of areas: e.g. the inclusion of more recent developments in LFG to account for discourse-pragmatic

1. For an overview of current theoretical approaches engaging with L2 developmental sequences, see *Language Learning Special Issue* (“Orders and sequences in L2 acquisition: 40 years on”) (2015, 65/1).

features (Pienemann et al. 2005), a model of the L2 initial mental grammatical system (Lenzing 2013, 2015), or the specification of developmental trajectories in a number of typologically diverse languages such as Swedish (e.g. Pienemann & Håkansson 1999; Håkansson 2005; Håkansson & Norrby 2007); German (e.g. Pienemann 1998; Jansen 2008) Japanese (e.g. Di Biase & Kawaguchi 2002; Kawaguchi 2005), Chinese (e.g. Zhang 2005, 2007; Gao 2004), Arabic (Mansouri 2005; Ghassan 2008) and Italian (Di Biase & Kawaguchi 2002; Di Biase 2008). Further developments include research on the acquisition of case (Baten 2013; Artoni & Magnani 2013), on L1 transfer (Lenzing et al. 2013; Pienemann et al. 2013; Håkansson et al. 2002), and on specific language impairment (Håkansson this volume), as well as studies on assessment/linguistic profiling (e.g. Keßler 2006), textbook analysis in terms of the learnability of grammatical structures (Lenzing 2008) and classroom research focusing on using tasks with a developmentally moderated focus on form to promote acquisition processes (Roos this volume).

Recent developments within the PT framework address both theoretical and applied issues, which reflects the continuous commitment to the application of research findings to language teaching and assessment. This objective is also reflected in the chapters in this volume. As in Hyltenstam and Pienemann's 1985 collection, this volume addresses not only current theoretical developments within the PT framework but also includes a section focussing on theory application.

With the expanded scope of research within the PT framework, different viewpoints on a number of theoretical and methodological issues have evolved. In terms of theoretical assumptions, these include, for instance, the exact relation between morphology and syntax in L2 acquisition and the status of grammatical functions in the L2 acquisition process. In relation to methodological considerations, multiple viewpoints exist concerning the exact application of the emergence criterion as well as the choice of different formats used in data elicitation. These in some ways controversial views are considered as a potential source of a continuous fruitful discussion on issues further research in PT needs to engage with. Some of these differences in perspective are also reflected in the papers in this volume. Despite the different opinions on these issues, these perspectives are united by the core assumptions of language processing and L2 development underlying PT, as well as the grammatical formalism of LFG. Reflecting the richness of debate within this field, the editors have not sought to impose theoretical agreement on contributors. Rather, we hope that readers find the different positions present in the volume stimulating for their own thinking and motivating for further research of their own.

References

- Atkinson, D. (Ed.). (2011). *Alternative approaches to second language acquisition*. New York, NY: Routledge.
- Artoni, D., & Magnani, M. (2013). LFG contributions in second language acquisition research: The development of case in Russian L2. In E. Butt & T. H. King (Eds.), *Proceedings of the LFG13 Conference* (pp. 69–89). Stanford, CA: CSLI.
- Baten, K. (2013). *The acquisition of the German case system by foreign language learners*. Amsterdam: John Benjamins. doi:10.1075/palart.2
- Clahsen, H. (1985). Profiling second language development: A procedure for assessing L2 proficiency. In Hyltenstam & Pienemann (Eds.), pp. 283–332.
- Clyne, M. G. (1985). Medium or object – different contexts of (school-based) second language acquisition. In Hyltenstam & Pienemann (Eds.), pp. 197–212.
- Di Biase, B., & Kawaguchi, S. (2002). Exploring the typological plausibility of Processability Theory: Language development in Italian second language and Japanese second language. *Second Language Research*, 18(3), 272–300.
- Fried, L. (1985). On the validity of second language tests. In Hyltenstam & Pienemann (Eds.), pp. 349–371.
- Gao, X. (2004). *Noun phrase morphemes and topic development in L2 Mandarin Chinese: A Processability Perspective*. Unpublished doctoral dissertation. Victoria University.
- Ghassan, A. S. (2008). The development of verbal structures in L2 Arabic. In J.-U. Keßler (Ed.), *Processability approaches to second language development and second language learning* (pp. 267–299). Newcastle upon Tyne: Cambridge Scholars.
- Håkansson, G. (2005). Similarities and differences in L1 and L2 development: Opening up the perspective: Including SLI. In M. Pienemann (Ed.), pp. 179–197. Amsterdam: John Benjamins. doi:10.1075/sibil.30.o8hak
- Håkansson, G., & Norrby, C. (2007). Processability Theory applied to written and oral Swedish. In F. Mansouri (Ed.), *Second language acquisition research: Theory-construction and testing* (pp. 81–94). Newcastle upon Tyne: Cambridge Scholars.
- Håkansson, G., Pienemann, M., & Sayehli, S. (2002). Transfer and typological proximity in the context of second language processing. *Second Language Research*, 18(3), 250–273. doi:10.1191/0267658302sr2060a
- Hulstijn, J.H. (1985). Testing second language proficiency with direct procedures. A comment on Ingram. In Hyltenstam & Pienemann, M. (Eds.), pp. 277–281.
- Hulstijn, J.H., Ellis, R., & Eskildsen, S.W. (2015). Orders and sequences in the acquisition of L2 morphosyntax, 40 years on: An introduction to the special issue. *Language Learning*, 65(1), 1–5. (Special issue *Orders and sequences in L2 acquisition: 40 years on*). doi:10.1111/lang.12097
- Hyltenstam, K. (1985). L2 learner's variable output and language teaching. In Hyltenstam & Pienemann (Eds.), pp. 113–136.
- Hyltenstam, K., & Pienemann, M. (Eds.). (1985). *Modelling and assessing second language acquisition*. Clevedon: Multilingual Matters.
- Hyltenstam, K., & Pienemann, M. (1985). Introduction. In Hyltenstam & Pienemann (Eds.), pp. 3–22.
- Ingram, D. E. (1985). Assessing proficiency: An overview on some aspects of testing. In Hyltenstam & Pienemann (Eds.), pp. 215–276.

- Jansen, L. (2008). Acquisition of German word order in tutored learners: A cross-sectional study in a wider theoretical context. *Language Learning*, 58(1), 185–231.
- Kawaguchi, S. (2005). Argument structure and syntactic development in Japanese as a second language. In M. Pienemann (Ed.), pp. 253–298. Amsterdam: John Benjamins. doi:10.1075/sibil.30.10kaw
- Kefler, J.-U. (2006). *Englischerwerb im Anfangsunterricht diagnostizieren. Linguistische Profilanalysen und der Übergang von der Primar- in die Sekundarstufe I.* (Giessener Beiträge zur Fremdsprachendidaktik). Tübingen: Gunter Narr.
- Lapkin, S. (1985). Pedagogical implications of direct second language testing: A Canadian example. In Hyltenstam & Pienemann (Eds.), pp. 333–347.
- Lenzing, A. (2008). Teachability and learnability: An analysis of primary school textbooks. In J.-U. Kefler (Ed.), *Processability approaches to second language development and second language learning* (pp. 221–241). Newcastle upon Tyne: Cambridge Scholars.
- Lenzing, A. (2013). *The development of the grammatical system in early second language acquisition. The Multiple Constraints Hypothesis.* Amsterdam: John Benjamins. doi:10.1075/palart.3
- Lenzing, A. (2015). Exploring regularities and dynamic systems in L2 development. *Language Learning*, 65(1), 89–122. (Special issue *Orders and sequences in L2 acquisition: 40 years on*). doi:10.1111/lang.12092
- Lenzing, A., Plessner, A., Hagenfeld, K., & Pienemann, M. (2013). Transfer at the initial state. *Zeitschrift für Anglistik und Amerikanistik. A Quarterly of Language, Literature and Culture*, 3, 265–286.
- Lightbown, P. (1985). Can language acquisition be altered by instruction? In Hyltenstam & Pienemann (Eds.), pp. 101–112.
- Long, M.H. (1985). A role for instruction in second language acquisition: Task-based language teaching. In Hyltenstam & Pienemann (Eds.), pp. 77–99.
- Mansouri, F. (2005). Agreement morphology in Arabic as a second language: Typological features and their processing implications. In M. Pienemann (Ed.), pp. 117–153. Amsterdam: John Benjamins. doi:10.1075/sibil.30.06man
- Nicholas, H. (1985). Learner variation and the teachability hypothesis. In Hyltenstam & Pienemann (Eds.), pp. 177–195.
- Ortega, L. (2009). *Understanding second language acquisition.* London: Hodder Education.
- Ortega, L. (2015). Second language learning explained? SLA across ten contemporary theories. In VanPatten & Williams (Eds.), pp. 245–272.
- Pienemann, M. (1985). Learnability and syllabus construction. In Hyltenstam & Pienemann (Eds.), pp. 23–77.
- Pienemann, M. (1998). *Language processing and second language development. Processability Theory.* Amsterdam: John Benjamins. doi:10.1075/sibil.15
- Pienemann, M. (Ed.). (2005). *Cross-linguistic aspects of Processability Theory.* Amsterdam: John Benjamins. doi:10.1075/sibil.30
- Pienemann, M. (2015). An outline of Processability Theory and its relationship to other approaches to SLA. *Language Learning*, 65, 123–151. (Special issue *Orders and sequences in L2 acquisition: 40 years on*). doi:10.1111/lang.12095
- Pienemann, M., & Håkansson, G. (1999). A unified approach towards the development of Swedish as L2: A processability account. *Studies in Second Language Acquisition*, 21(3), 383–420. doi:10.1017/S0272263199003022

- Pienemann, M., Di Biase, B., & Kawaguchi, S. (2005). Extending Processability Theory. In M. Pienemann (Ed.), pp. 199–251. Amsterdam: John Benjamins. doi:10.1075/sibil.30.09pie
- Pienemann, & Kessler, J.-U. (Eds.). (2011). *Studying Processability Theory*. Amsterdam: John Benjamins. doi:10.1075/palart.1
- Pienemann, M., Kessler, J.-U., & Lenzing, A. (2013). Developmentally moderated transfer and the role of the L2 in L3 acquisition. In A. Mattsson & C. Norrby (Eds.), *Language acquisition and use in multilingual contexts: Theory and practice* (Travaux de l'Institut de linguistique de Lund 52) (pp. 142–159). Lund: University of Lund.
- Pienemann, M., & Lenzing, A. (2015). Processability Theory. In VanPatten & Williams (Eds.), pp.159–179.
- Stölting, W. (1985). Language assessment as social activity. In Hyltenstam & Pienemann (Eds.), pp. 379–395.
- VanPatten, B., & Williams, J. (Eds.). (2015). *Theories in second language acquisition: An introduction* (2nd ed.). New York, NY: Routledge.
- Zhang, Y. (2005). Processing and formal instruction in the L2 acquisition of five Chinese grammatical morphemes. In M. Pienemann (Ed.), *Cross-linguistic aspects of Processability Theory* (pp.155–177). Amsterdam: John Benjamins. doi:10.1075/sibil.30.07zha
- Zhang, Y. (2007). Testing the topic hypothesis: The L2 acquisition of Chinese syntax. In F. Mansouri (Ed.), *Second language acquisition research: Theory-construction and testing* (pp.137–163). Newcastle upon Tyne: Cambridge Scholars.

About this book

The book at hand is the fifth volume of the PALART series. It is divided into two major parts. While the first part, called *Theory Development*, engages with a number of aspects related to theoretical developments within the framework of PT, the second one named *Theory Application* investigates approaches on the assessment of second languages. As the titles of the two sections suggest, both major foci of this volume go together hand in hand: learners need to develop their second (or any other than their first) language, and teachers, instructors as well as SLA researchers need knowledge about theory and assessment of second languages.

The chapters in the first part address a number of crucial issues in SLA research, such as the question of the nature of the L2 initial state, the relationship between vocabulary and syntax, the issue of second language impairment and the role of transfer in L2 acquisition.

All chapters present research on second language development within the framework of Processability Theory (cf. Pienemann 1998, 2005; Pienemann & Keßler 2011, 2012; Pienemann & Lenzing 2015) and illustrate the wide range of PT-based research on SLA:

In the first chapter, Lenzing focuses on the initial state in L2 acquisition and, in particular, on the development of argument structure in the mental grammatical system of early L2 learners. She proposes specific constraints at the semantic and syntactic level of linguistic representation in the L2 initial state. Her hypotheses are formalised in the *Multiple Constraints Hypothesis* (Lenzing 2013), a model of the initial L2 mental grammatical system that constitutes a conceptual extension of PT. Supporting evidence for her claims concerning the development of argument structure comes from a combined cross-sectional and longitudinal study of L2 learners of English with German as L1 in a primary school context.

The second chapter by *Kawaguchi* investigates the relationship between vocabulary size and syntactic development in L2 acquisition within the framework of PT. She presents the results of a cross-sectional study with L2 learners of English with Japanese as L1. The focus of the study is on the development of question formation and constructions that require both linear and non-linear argument function mapping by learners with different levels of vocabulary knowledge. The results show a correlation between the learners' vocabulary size and their development in question formation and non-linear argument-function mapping.

Another important issue widening the scope of PT, namely the language of children suffering from SLI, is investigated by *Håkansson* in the following chapter. SLI characteristics differ cross-linguistically, which leads to seemingly contradictory research findings. Taking a developmental perspective instead of assuming representational deficits, *Håkansson* provides an explanation for the contradictions. She presents a study on Swedish children with SLI and demonstrates how their language development can be explained in terms of PT by looking at the children's language development individually and analysing them as language learners at different developmental stages.

The fourth chapter by *Pienemann, Lenzing & Keßler* engages with the ongoing debate in SLA research about the role of transfer in L2/L3 acquisition. Viewing transfer within the framework of the *Developmentally Moderated Transfer Hypothesis*, the authors critically review the claim that the L3 initial word order is determined by the L2 and identify a number of theoretical and methodological weaknesses of previous studies supporting this claim. They present a study on the acquisition of Swedish as L3 by adult German L1 speakers with different L2s. The results support the claim that learners only transfer structures when they are developmentally ready to process the features to be transferred.

As mentioned in the first lines of this introduction, the second part of this book focuses on the assessment of second languages. Obviously, in language acquisition there are various competences to be acquired by L2 learners. All chapters of this second part deal with important aspects of second language assessment. The authors take close looks at relevant competences, different age groups and also complementary approaches to language assessment and how those might even benefit from each other.

Zhang and Liu address the widely-debated question of why Chinese learners of English show high variability in the acquisition of the past -ed marker. They hypothesise that potential reasons for this phenomenon are that the variability in past tense marking (1) reflects the learners' university training and (2) is related to the *Bad Choice Hypothesis*. The results of a study of highly advanced Chinese L1 speakers of English indicate that high-quality training programmes led to a higher attainment of the past -ed marker and seemed to discourage bad choices by the learners in other domains of morphology.

Roos addresses the questions of *what* should be taught *when* and *how* in the foreign language classroom. In particular, she explores the potential of communicative tasks with a developmentally moderated focus on form in promoting the L2 acquisition process. In her chapter, she provides an exemplary discussion of the use of two sets of communicative tasks, one focussing on the 'plural -s' and on the 'third person singular -s'. She shows that the use of tasks with a developmentally

moderated focus on form in the second language classroom has the potential to facilitate and enhance the L2 acquisition process.

Hagenfeld presents a pilot study that investigates possible interfaces between psychometric rating scales based on the Common European Framework of Reference (CEFR) and linguistic profiling and investigates whether and to what extent the PT based diagnostic tools *Rapid Profile* and *Auto Profile* can be integrated into proficiency rating. In this way, shortcomings of the CEFR, such as the lack of precision due to its broad scope, can be addressed. Her results indicate a correlation between CEFR level and PT stages. This is particularly the case for beginning learners at low levels of the CEFR.

Maier, Neubauer, Schwirz, Couve de Murville and Kersten investigate selected primary school concepts of L2 learning. Effects of immersion programmes, their potential and the chances and challenges of linguistic profiling for advanced learners are given a major interest in this chapter. The authors analyse various learners who have been taught in immersion programs and traditional teaching programs. The analysis shows which developmental stages learners reach and which factors might influence the results. In addition, they discuss whether the PT stages and communicative tasks can be used for assessment.

Keßler and Liebner present a task-based approach to an L2 diagnosis with PT and *Rapid Profile*. They apply their diagnostic approach to a teaching unit for intermediate learners and offer an example unit, which combines task-, literature- and media-based lessons. The idea of using Podcasts within the unit helps to collect language data of whole language classes in a school setting. They also show how it can easily be adapted to various other units in a language classroom and thereby demonstrate the potential of *Rapid Profile* as a diagnostic tool in a school setting.

In the closing chapter, *Rossa* takes a close look at the validity of an EFL listening comprehension test that was developed for a large-scale assessment project. He analyses 18 language learners from a German secondary school using a think-aloud-technique to detect construct-relevant and irrelevant processes involved in L2 listening comprehension tests. *Rossa's* findings show that the chosen think-aloud technique gives insight into the participants' cognitive process while working on the task and that the test is successful in its construct validity.

The two complementary foci of this volume, namely the development and the assessment of language acquisition and learning, are investigated from various perspectives. The volume contributes to a better understanding of how languages are acquired and indicates possibilities to assess language acquisition. The book is therefore helpful and important for various groups involved in researching, teaching and learning foreign languages, e.g. SLA researchers, teacher trainers, teacher trainees, teachers and advanced students in various SLA and linguistic programmes.

References

- Lenzing, A. (2013). *The development of the grammatical system in early second language acquisition: The Multiple Constraints Hypothesis*. Amsterdam: John Benjamins. doi:10.1075/palart.3
- Pienemann, M. (1998). *Language processing and second language development: Processability Theory*. Amsterdam: John Benjamins. doi:10.1075/sibil.15
- Pienemann, M. (Ed.). (2005). *Cross-linguistic aspects of Processability Theory*. Amsterdam: John Benjamins. doi:10.1075/sibil.30
- Pienemann, M., & J.-U. Kefßler (Eds.). (2011). *Studying Processability Theory. An introductory textbook*. Amsterdam: John Benjamins. doi:10.1075/palart.1
- Pienemann, M., & Kefßler, J.-U. (2012). Processability Theory. In S. Gass & A. Mackey (Eds.), *Handbook of second language acquisition* (pp. 228–247). New York, NY: Routledge.
- Pienemann, M., & Lenzing, A. (2015). Processability Theory. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition. An introduction* (2nd ed.) (pp. 159–179). New York, NY: Routledge.

PART I

Theory Development

The development of argument structure in the initial L2 mental grammatical system

Anke Lenzing

Paderborn University

This chapter investigates the development of argument structure in early L2 acquisition. I view argument structure and its development within the context of the *Multiple Constraints Hypothesis* (Lenzing 2013) and its core claim that the L2 initial mental grammatical system is constrained at the different levels of linguistic representation. I argue that at the beginning of the L2 acquisition process, argument structure is not fully developed. In particular, I claim that essential syntactic features are missing which are required to align semantic and syntactic information in the L2 speech production process. The constraints on argument structure lead to direct mapping processes from arguments to surface form. I present a combined cross-sectional and longitudinal study of beginning learners of L2 English with German as L1 in a formal context. The analysis of the oral speech production data focuses on argument structure and its development in the L2 acquisition process. The results of the analysis support my claims concerning the initial constraints at the level of argument structure.

1. Introduction

This paper focuses on the development of argument structure (a-structure) in the initial L2 mental grammatical system of beginning learners of English with German as L1 in a formal context. I claim that a-structure is not fully developed at the beginning of the L2 acquisition process. I hypothesise that essential features are missing at a-structure level that are required to align semantic and syntactic information in the L2 speech production process.

The view on L2 a-structure adopted in this paper is based on the *Multiple Constraints Hypothesis* (MCH) (Lenzing 2013). The MCH is situated in the theoretical framework of *Lexical-Functional Grammar* (LFG) (Bresnan 2001) and *Processability Theory* (PT) (Pienemann 1998; Pienemann et al. 2005). Its core claim is that the L2 initial mental grammatical system is not fully developed in terms of mental representations. I hypothesise that the initial L2 mental grammatical system is highly constrained at the different levels of linguistic representation

spelled out in LFG and that these restrictions also apply to the level of a-structure. The initial restrictions at L2 a-structure level result in the learners' inability to map arguments onto grammatical functions. I argue that beginning L2 learners rely on direct mapping processes from arguments onto surface form. The overall development of the grammatical system of early L2 learners is in line with PT and can be explained in terms of both feature unification and mapping processes.

In this paper, I will outline the theoretical basis for my claims and then present the results of a combined cross-sectional and longitudinal study of the oral speech production data of early L2 learners of English that focuses on L2 a-structure and its development.

In a first step, I will introduce the notion of a-structure as conceptualised in LFG as well as the principles that guide the mapping processes from arguments to grammatical functions. This is followed by a brief outline of the development of mapping operations in the course of SLA according to PT. In a next step, the basic premises of the MCH are presented with particular focus on the constraints at a-structure level. Then I provide an overview of the study and its research design. In the following, the actual analysis of a-structure is presented and in the final part, the results of the analysis are discussed.

2. Argument structure in LFG

The question as to what kind of linguistic representation of a-structure the L2 learner can make recourse to at the L2 initial state is naturally related to the question of how a-structure is conceptualised in a fully developed mental grammatical system. In order to get a complete picture of the constraints at the level of a-structure proposed in this paper, it is important to gain insights into the full representation of a-structure as well as the mapping principles guiding the process of the alignment of semantic and syntactic information in LFG.

A central component of LFG is its projection architecture with three independent levels of linguistic representation that exist in parallel and are related to each other by specific linking or mapping principles. The three levels are functional structure (f-structure), constituent structure (c-structure) and argument structure (a-structure). In f-structure, universal aspects of grammar are encoded; it contains grammatical functions, such as subject or object. The second level that represents syntactic concepts is c-structure. It is at this level that the surface syntactic organisation of phrases is represented (cf. Dalrymple 2001:45), i.e. the structural relations between the words that make up a sentence are depicted in terms of phrase structure trees. In contrast to f-structure, c-structure is language-specific. As the main focus of this paper is on the development of a-structure, this concept will be explained in more detail.

of patient ('the boy'). In Example (1), the locative role is represented by (z) ('on the table'), which is ordered to the right of (y), as the locative is the least prominent role in the thematic hierarchy. Finally, the most prominent role of the predicator 'freeze' is represented by (x) and takes the role of theme ('Mary') (cf. Bresnan 2001: 307).

As for the semantic side of a-structure, all (x) arguments represent the most prominent semantic roles of the respective predicators. However, these arguments differ in terms of their syntactic properties. These syntactic differences of the (x) arguments are captured by the syntactic side of a-structure, specifically by the syntactic features encoded in a-structure. These syntactic features constrain the mapping process of the thematic roles in a-structure onto the argument functions in f-structure.

The mapping principles from a-structure to f-structure are spelled out in detail in *Lexical Mapping Theory* (Bresnan & Kanerva 1989; Bresnan 2001). A core idea underlying *Lexical Mapping Theory* is that certain thematic roles are restricted as to the grammatical functions they can be mapped onto and that certain grammatical functions can only be filled by a restricted type of thematic roles. This observation led to the classification of the basic argument functions SUBJ, OBJ, OBL_θ and OBJ_θ according to the features [\pm -r] (thematically unrestricted or not) and [\pm o] (objective or not):

(5) Feature Decomposition of Argument Functions

	-r	+r
-o	subj	obl _θ
+o	obj	obj _θ

(taken from Bresnan 2001: 308)

The features [\pm r] and [\pm o] indicate whether a syntactic function is restricted in terms of its thematic role. Both SUBJ and OBJ function are not restricted as regards the thematic role they can take and therefore, they are classified as [\pm r]. However, this is not the case for OBL_θ and OBJ_θ. These two functions are restricted to specific thematic roles and are therefore classified as [\pm r]. The features [\pm o] and [-o] refer to objective and non-objective syntactic functions. OBJ and OBJ_θ are both object functions and are therefore classified as [\pm o]. As SUBJ and OBL_θ are not object-type functions, they are classified as [-o].

The question of whether a-structure is universal or whether it also exhibits language-specific aspects is not explicitly resolved in the LFG literature. In Lenzing (2013), I hypothesise that a-structure contains both universal and language-specific components. In particular, I argue that the argument roles themselves, their hierarchical ordering and their syntactic classification are universal, whereas the actual arguments that the respective predicators take are language-specific.

The three levels of representation outlined above do not only model different aspects of grammar and exhibit specific properties; in keeping with the projectional architecture of LFG, they are furthermore related to each other by specific mapping principles (Bresnan 2001: 20). As the focus of this paper is the development of a-structure, the mapping principles underlying a-to f-structure mapping as conceptualised in *Lexical Mapping Theory* are of particular relevance and are therefore briefly outlined below.

Following Bresnan & Kanerva (1989), there are three lexical mapping principles that relate the thematic roles encoded in the semantic side of a-structure to the syntactic features in the syntactic side:

- a. Intrinsic role classifications
- b. Morpholexical operations
- c. Default classifications

Firstly, intrinsic role classifications relate the intrinsic properties of thematic roles to specific syntactic functions. The *agent encoding principle* states that the intrinsic value of the role agent is constrained to [-o]. The *theme encoding principle* constrains the intrinsic value of the patient/theme role to [-r] which results in the patient/theme being realised as either subject or object. The third principle, the *locative encoding principle*, ensures that the locative receives the feature [-o] and is realised as subject or oblique. These classifications are considered to be universal and therefore they apply cross-linguistically (cf. Bresnan & Kanerva 1989: 26).

Secondly, morpholexical operations add or suppress thematic roles in lexical argument structure. This is for instance the case in the passive, where a morpholexical operation leads to the suppression of the logical subject (i.e. the agent), so that the unrestricted patient is mapped onto the SUBJ function instead. This is illustrated in Figure 1 below:

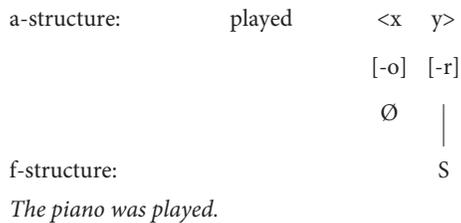


Figure 1. Morpholexical operations in passives

In a final step, default classifications apply once the argument structure has been built up in a morpholexical fashion. These classifications ensure that the highest thematic role is assigned the SUBJ function and that all other roles that are

lower in the hierarchy are assigned non-subject functions (cf. Bresnan & Kanerva 1989:27). It should be noted at this point that, “all default classifications apply to a role only if it is not already specified for an incompatible value of the default feature” (Bresnan & Kanerva 1989:28).

The following two wellformedness conditions on lexical form further constrain the mapping process from a- to f-structure:

Function-Argument Bi-uniqueness:

Each a-structure role must be associated with a unique function, and conversely.

The Subject Condition:

Every predicator must have a subject. (Bresnan 2001:311)

The instantiation of both mapping principles and wellformedness conditions is illustrated below with the example of ‘place’ as in ‘John placed the plate on the table’.

place	<	x	y	z	>	
		(agent)	(theme)	(locative)		
		[-o]	[-r]	[-o]		intrinsic role classification
		[-r]		[+r]		default classification
		SUBJ	SUBJ/OBJ	OBL ₀		
			OBJ			Function/arg. biuniqueness
		John placed the plate		on the table		

Figure 2. Principles and constraints in a- to f-structure mapping

As shown in Figure 2, the verb ‘place’ takes three arguments. In a first step, the *intrinsic role classification* assigns the agent (‘John’) the feature [-o]. The theme (‘the plate’) is classified as [-r] and the locative (‘on the table’) as [-o]. In a next step, the *default classification* assigns the feature [-r] to the agent role and the feature [+r] to the locative. In a final step, the *Function-Argument Bi-uniqueness* condition applies which specifies that the theme is mapped onto the OBJ function.

In sum, the principles of *Lexical Mapping Theory* specify the selection of grammatical functions in f-structure on the basis of the classification of the arguments in a-structure. In this way, the theory accounts for the mapping process from a-structure to f-structure in a precisely defined way.

After having summarised the core premises of argument structure and *Lexical Mapping Theory* in LFG, I now turn to a brief overview of the mapping processes in PT.

3. Mapping processes in PT

The mapping principles between the different levels of linguistic representation as conceptualised in LFG were incorporated in the extended version of PT in order to account for a range of discourse-pragmatic structures and exceptional verbs, which are characterised by their underlying linguistic nonlinearity. The relationship between a-structure, f-structure and c-structure is not necessarily linear, as there is considerable surface structure variation. Therefore, different mapping principles account for instance for differences between active and passive, or affirmative sentences and question forms (cf. Pienemann et al. 2005: 201).

With regard to the developing L2 system, the core hypothesis of the extended version of PT (Pienemann et al. 2005) is that learners begin with *unmarked alignment*, i.e. linear default correspondences between a- f- and c-structure (cf. Pienemann & Lenzen 2015: 168) (see Figure 3).

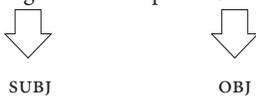
Mapping process	Structures	Example
Linear default mapping	a-structure	<i>play</i> < agent patient/theme >
	f-structure	
	c-structure	
		<i>John</i> played <i>the guitar</i> NP _{subj} NP _{obj}

Figure 3. Linear correspondence relationship between the three levels of representation (Lenzen (2013:94), based on Pienemann et al. 2005)

This is captured in the *Unmarked Alignment Hypothesis* (UAH) which predicts that

[i]n second language acquisition learners will initially organise syntax by mapping the most prominent semantic role available onto the subject (i.e. the most prominent grammatical role). The structural expression of the subject, in turn, will occupy the most prominent linear position in c-structure, namely the initial position. (Pienemann et al. 2005: 229)

In the course of L2 development, the learners acquire additional processing resources that enable them to process more complex linguistic structures that are characterised by non-linear correspondences either between a-structure and f-structure (e.g. the passive) or c-structure and f-structure (e.g. object topicalisation).

A deviation from the UAH that creates linguistic non-linearity is the occurrence of non-subjects, such as adverbials and Wh-words, in sentence-initial position. The preposing of adjuncts to canonical structure and certain discourse functions such as Focus and Topic require the assignment of one of the non-argument functions Topic, Focus or Adjunct to the constituents that are adjoined to XP (Pienemann et al. 2005:232). Here, the mapping process between c-structure and f-structure is no longer linear, as in the sentence ‘Yesterday everyone smiled’ or in the question ‘What did he buy?’. In these cases, the subject no longer occurs in sentence-initial position (cf. Pienemann et al. 2005:236). This is illustrated in Figure 4.

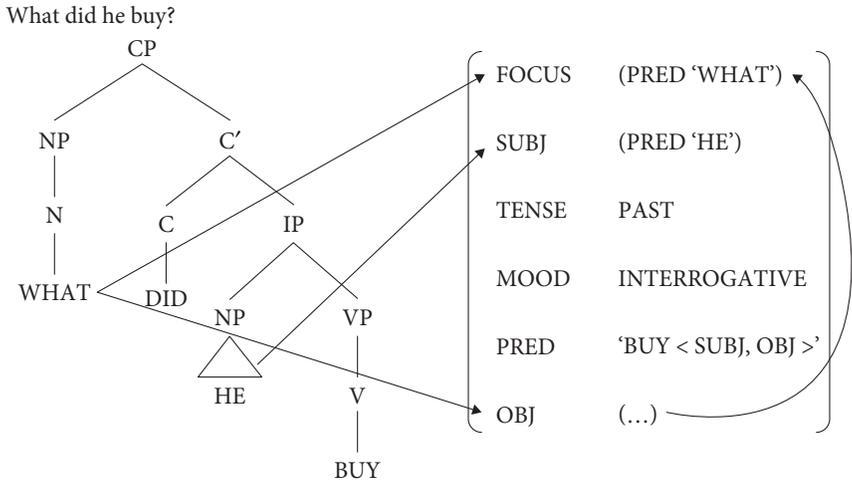


Figure 4. Non-linear mapping in Wh-questions (adapted from Pienemann et al. 2005:211)

In the Wh-question ‘What did he buy?’, the linguistic non-linearity is created by the fact that the Wh-word occurs in initial position and is mapped onto both the object and the focus function.

The acquisition of the mapping processes from c- to f-structure is captured in the TOPIC hypothesis which states that

[i]n second language acquisition learners will initially not differentiate between SUBJ and TOP. The addition of an XP to a canonical string will trigger a differentiation of TOP and SUBJ which first extends to non-arguments and successively to [core] arguments thus causing further structural consequences.

(Pienemann et al. 2005:239)

As for the relation between a-structure and f-structure, the principles of *Lexical Mapping Theory* have been incorporated in the PT framework. The acquisition of the non-linear correspondences between a- and f-structure is spelled out in the *Lexical Mapping Hypothesis* which is briefly outlined in the following.

Similar to the correspondences between c- and f-structure, the relation between a- and f-structure undergoes certain changes in the course of L2 development. A first deviation from the linear mapping process from a- to f-structure is the passive. As was explained above, in the case of the passive, the suppression of the logical subject in a-structure leads to the mapping of the patient/theme onto the subject function. The agent can optionally be realised as OBL_{ag} (see Figure 5).

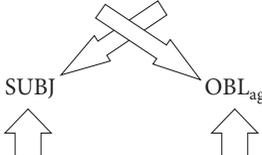
a- to f- structure mapping	Structures	Example
Non-default mapping. (single clause) passive	a-structure	<i>play</i> < agent patient/theme >
	f-structure	
	c-structure	<i>The <u>guitar</u> was played by <u>John</u>.</i>

Figure 5. Non-default mapping in passive construction (Lenzing (2013: 103), based on Pienemann et al. 2005)

A more complex form of non-linear correspondences between a- and f-structure is the case of causative constructions. The linguistic non-linearity in causative constructions is due their intrinsic a-structure which results in the fusion of two arguments onto one grammatical function (cf. Alsina 1996:193; Pienemann et al. 2005: 244).

The following table depicts the mapping processes from a- to f-structure that account for the different structural outcomes at the different levels of development.

a- to f-structure mapping	structural outcomes
Non-default, complex mapping	Complex predicates e.g. causative (in Romance languages, Japanese, etc.), raising, light verbs
↑ Non-default mapping. (single clause)	↑ Passive Exceptional verbs
↑ Default mapping, i.e. Most prominent role is mapped onto subject	↑ Canonical Order

Figure 6. Lexical Mapping Hypothesis (taken from Pienemann et al. 2005: 240)

With this understanding of the mapping processes underlying PT, I now address some of the key premises of the *Multiple Constraints Hypothesis* with particular focus on the constraints at a-structure level in the L2 initial mental grammatical system.

4. The Multiple Constraints Hypothesis

As mentioned in the introduction, the hypotheses concerning the constraints on argument structure and its development form part of the *Multiple Constraints Hypothesis* (MCH) (Lenzing 2013), a theoretically-motivated model of the initial L2 mental grammatical system. The MCH is illustrated in Figure 7 and briefly summarised below:

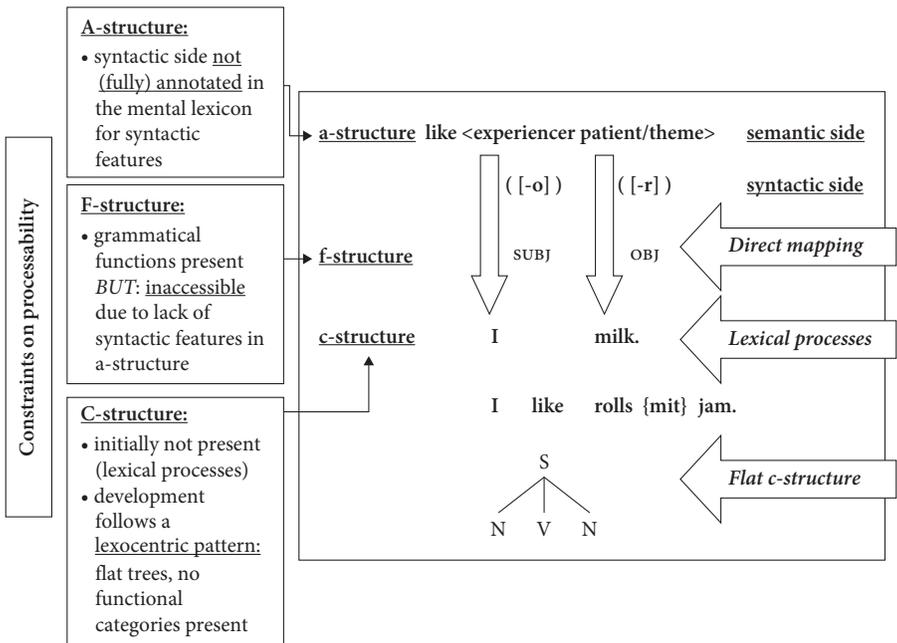


Figure 7. The Multiple Constraints Hypothesis (Lenzing 2013:8)

The core claims underlying the MCH are that the three levels of linguistic representation in LFG are not fully developed in terms of mental representations in the L2 initial state and that only a restricted set of formal linguistic categories is initially present.

A central idea underlying the MCH is that the lexicon is being successively annotated. This implies that at the beginning of L2 acquisition, not all lexical items are annotated for their syntactic category (e.g. noun, verb). The successive annotation of the L2 lexicon also includes the verbs' arguments: I hypothesise that not all verbs are annotated for both number and type of arguments they take (e.g. agent, patient). The incomplete annotation of the L2 lexical items results in utterances such as *'It's a pink?' (learner 06). Here, the adjective pink occurs in the wrong position in the sentence, as it is not annotated for its syntactic category 'adjective'.

I argue that in the L2 initial grammatical system, no c-structure is present so that learners rely on lexical processes at the beginning of their L2 acquisition process. The gradual development of c-structure follows the predictions spelled out in PT: it is characterised by a development from basic, flat-c-structures to more complex, hierarchical ones.

The MCH makes the assumption that the universal grammatical functions (SUBJ, OBJ etc.) at f-structure level are present in the L2 initial state. Additionally, I hypothesise that initially, the grammatical functions are inaccessible, as the mapping process from a- to f-structure is blocked due to missing features at a-structure level. This results in direct mapping processes from a- to f-structure, i.e. from arguments to surface form.

As for the level of argument structure, I assume that a restricted set of argument roles is present in the L2 initial state. The constraints at this level mainly concern the syntactic side of a-structure: I hypothesise in the MCH that a-structure is initially not annotated for its syntactic features (i.e. +r/-r and +o/-o), which are essential to map arguments onto grammatical functions at f-structure level (see Section 3 above). The lack of syntactic features at a-structure level results in the inability to map arguments onto grammatical functions at f-structure. In line with this, I argue that in the L2 initial state, learners rely on direct mapping operations from arguments to surface form without recourse to the grammatical functions at f-structure level.

This constitutes a modification of the *Unmarked Alignment Hypothesis*: in contrast to the claim that the initial mapping process is characterised by one-to-one correspondences between the three levels of linguistic representation (a-structure, f-structure and c-structure), the MCH proposes that at the beginning of the L2 acquisition process, the arguments are mapped directly onto surface structure due to the incomplete annotation of the syntactic side of a-structure. This direct mapping process is illustrated in Figure 8.

I now turn to a study of early learner language to test my claims concerning the constraints at the level of a-structure in the L2 initial mental grammatical system.

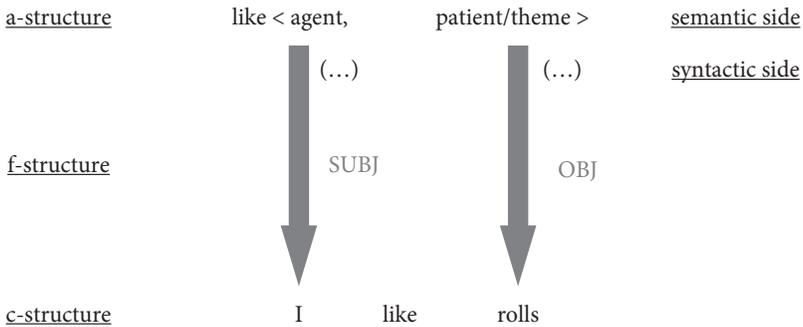


Figure 8. Direct mapping in early L2 acquisition (Lenzing 2013:222)

5. The study & research design

The study I present here forms part of a larger study on early L2 acquisition presented in Lenzing (2013). The combined cross-sectional and longitudinal study investigates the oral speech production of L2 English by 24 beginning learners with German as L1 in a primary school context. The data collection was carried out at four different primary schools in and around Paderborn, Germany and the participants were 24 students in total (six students per school). The elicitation of oral speech production data took place at two points in time: the students were interviewed at the end of grade 3 (cf. Roos 2007) and at the end of grade 4, i.e. after one and after two years of formal instruction in English. In line with the curriculum, the students received two English lessons per week. For the purpose of data elicitation, six different communicative tasks were used which were based on the vocabulary of the textbook and the lessons.² The tasks were designed specifically to provide a context for spontaneous oral speech production. Moreover, they aimed at specific syntactic and morphological structures, such as question forms or the third person singular -s, in order to establish a profile of the interlanguage grammar of each learner and to determine the individual developmental stages.³ In all four classes, the textbook *Playway* (Gerngross & Puchta 2003a, 2003b) served as the

2. For each round of data elicitation, three communicative tasks were used. The tasks in the first round were designed by Roos (2007), the tasks in the second round of elicitation were designed by the author (see Lenzing 2013:146ff.)

3. For an overview of the role of communicative tasks in data elicitation, see e.g. Pienemann (1998), Pienemann & Mackey (1993); Mackey & Gass (2005), Gass & Mackey (2007).

basis for the lessons. Therefore, the thematic units covered in the lessons were largely identical across the different classes. The communicative tasks were based on the respective textbooks which had the advantage that the pupils were familiar with the vocabulary of the tasks. Furthermore, it allowed for a reliable comparison of the individual learners' development. In order to reduce anxiety, the pupils were interviewed in pairs (cf. Roos 2007; Johnstone 2000). The recordings took place at the respective schools and the individual recording sessions lasted between 15–25 minutes. The data were audiotaped, transcribed and analysed according to the criteria outlined in the following.

6. Analysis

The analysis presented here focuses on two aspects. Firstly, a linguistic profile analysis of the learners' speech samples was carried out and the developmental stages of the individual learners according to the PT hierarchy were determined. Secondly, a distributional analysis of the a-structure of the lexical verbs occurring in the learners' utterances was conducted.

In the analysis of a-structure, the utterances were classified according to the following four categories (see Lenzing 2013: 212f.):

a. Formulaic sequences

Those utterances that occur invariantly in the learners' speech are classified as formulaic sequences. Within this category, a further distinction is made between *formulae*, i.e. sequences which are introduced as fixed expressions in the learners' textbooks and *formulaic patterns*, i.e. sequences consisting of an unanalysed chunk and an open slot that can be filled with different lexical material. The status of these expressions as formulaic sequences was determined by means of a distributional analysis. In the analysis presented here, it is claimed that these units are memorised as chunks and stored holistically by the learners. In line with this, it is hypothesised that, in these cases, no complete a-structure is present as the verb is not stored as a separate lexical entry.

b. Translation (grade 4 only)

In the current context, the term *translation* means that the lexical verb that occurs in the utterance had been previously translated, i.e. the learner did not know the L2 word and asked the interviewer for a translation. Furthermore, it is hypothesised that, in utterances with translated verbs, both the verb and its a-structure are still annotated in the learner's first language.

c. Non target-like argument structure:

The a-structure of the verb is considered to deviate from the target-like pattern in the following five cases:

1. when one or more arguments are expressed in the learner's mother tongue (German)
2. when one or more arguments are missing in the utterance
3. when there are too many arguments in the utterance
4. when the arguments that are expressed by the learner are not the intended ones
5. when the entry of the verb is not fully annotated in the lexicon.

d. Target-like argument structure

An a-structure is considered to be target-like if the verb is expressed with its corresponding arguments. The syntactic structure is not relevant in this context.

As outlined in Section 4, one key assumption concerning the constraints at the level of L2 a-structure is that its syntactic side lacks essential features that are required for a-to f-structure mapping, so that learners rely on direct mapping processes from arguments to surface structure.

In keeping with these hypotheses, I assume that initially, the learner utterances display more target-like a-structures in statements than in question forms. This is due to the fact that statements display canonical word order. In this case of default mapping, the arguments do not necessarily need to be annotated for syntactic features as the canonical word order of statements allows for a direct mapping from argument onto surface form (see Figure 8). However, following the Topic Hypothesis (Pienemann et al. 2005), question forms constitute a departure from the direct mapping process due to the underlying non-linearity (see Section 3). Although the non-linearity that is present in question forms is created by the non-default mapping between c-structure and f-structure, it is argued here that in order for the required mapping principles to be applied, the arguments need to be specified for their grammatical functions.

In line with these considerations, the focus of the actual analysis of a-structure is the following: First of all, a distinction is made between a-structure in questions and in statements. This distinction is motivated by the hypothesis outlined above that the learner utterances contain more non target-like a-structures in statements than in questions due to the lack of syntactic features in a-structure. In a second step, the differences between a-structure after one and after two years of instruction are outlined. This is done in order to account for the development of a-structure in the grammatical system of the early L2 learners.

7. Results

The main results of the analysis are presented according to the following sequence. Firstly, the findings for the speech samples of grade 3 learners are given and discussed. In a second step, the results of the analysis of the grade 4 learners are presented. Finally, the results of the two learner groups will be compared.

7.1 Grade 3 – Developmental stages

To diagnose the development of the second language learners and to determine the stage of development in their L2 according to the processability hierarchy, a linguistic profile of the individual learners was created by carrying out a distributional analysis of the relevant syntactic and morphological features in the learners' speech samples.⁴

Table 1 provides an overview of the developmental stages of the grade 3 learners.

Table 1. Overview of developmental stages of grade 3 learners (adapted and modified from Roos 2007: 164)⁵

Stage	Group 1						Group 2					
	C01	C02	C03	C04	C05	C06	C07	C08	C09	C10	C11	C12
6	-	-	-	-	-	-	-	-	-	-	-	-
5	-	-	-	-	-	-	-	-	-	-	-	-
4	-	-	-	-	-	-	-	-	-	-	-	-
3	-	-	-	-	-	-	-	-	-	-	(+)	-
2	-	-	-	+	-	-	-	-	-	-	+	-
1	+	+	+	+	+	+	+	+	+	+	+	+

4. For a detailed discussion of the results, see Roos (2007).

5. Apart from two learners, C18 and C21, who moved within the school year, the same learners were interviewed after one and after two years of instruction. These two particular learners are labelled as C18.1/C18.2 and C21.1/C21.2 respectively.

Stage	Group 3						Group 4					
	C13	C14	C15	C16	C17	C18.1	C19	C20	C21.1	C22	C23	C24
6	-	-	-	-	-	-	-	-	-	-	-	-
5	-	-	-	-	-	-	-	-	-	-	-	-
4	-	-	-	-	-	-	-	-	-	-	-	-
3	-	-	-	-	-	-	-	-	-	-	-	-
2	-	-	-	-	-	-	-	-	-	-	-	-
1	+	+	+	+	+	+	+	+	+	+	+	+

As can be seen in Table 1, the majority of the learners (22) are at stage 1 after one year of instruction in English. This means that their productive utterances are restricted to single words, idiosyncratic utterances⁶ and formulaic sequences (cf. Roos 2007; Lenzing 2013). Two learners (C04, C11) produce SVO-structures (stage 2) and only one of them (C11) starts to use some stage 3 features.

7.2 Argument structure grade 3 – questions

In the analysis of argument structure, in a first step the lexical verbs that occur in question forms in the learners' speech at the end of grade 3 were determined. Table 2 shows that the learners produce three types of lexical verbs (the number of different verbs occurring in the data) and a total of eight tokens (the total number of verbs occurring in the data). Their question forms contain only verbs that take two arguments, such as 'like'. This means that intransitive or ditransitive verbs do not occur at all. The thematic roles that the arguments take are restricted to agent, experiencer and patient/theme. None of the verbs has previously been translated.⁷

6. Idiosyncratic utterances are those learner utterances that are semantically and syntactically ill-formed so that the meaning can only be inferred from the context, such as the question form 'Do you I am animal?' (=Do you have an animal?) (learner C15) (See Lenzing 2013:171ff.).

7. As pointed out by Lenzing (2013:214), the limited number of utterances containing lexical verbs produced by the individual learners is due to the fact that these learners are at the very beginning of their acquisition process.

Table 2. Lexical verbs in question forms – types & token (Lenzing 2013:214)

Verb types	Verb tokens	Number of arguments	Arguments – thematic roles
3	8	2	agent, experiencer, patient/theme

In a second step, I analysed the actual occurrence of question forms with lexical verbs in the learners' speech with regard to their argument structure. This analysis is presented in Table 3 below.

Table 3. Distributional analysis of a-structure – grade 3 learners (Lenzing 2013:215)

	C01	C06	C08	C11	C22	C24
Formulae occurring in textbooks	1	1		1	1	
Non target-like a-structure:						
Argument missing	1					
Too many arguments			1			
Argument not the intended one		1				
Unclear argument structure						1

Table 3 reveals that merely six learners produced utterances with lexical verbs in questions. The question forms consist either of formulae in textbooks, or they are ill-formed in terms of the arguments the learner expresses (see examples below).

To summarise the results, half of the questions the learners produce (50%) are classified as formulaic. In the remaining 50% of the utterances, the a-structure deviates from the target-like pattern. This means that none of the questions that occur in the learners' speech sample displays a fully productive target-like a-structure (see Lenzing 2013:215).

The questions that are classified as formulaic consist of formulae occurring in the learners' textbook (cf. (6)).

(6) C01 What do you like for breakfast?

This question form can be unambiguously assigned to a specific unit (*Unit 8 Breakfast*) in the textbook *Playway 3* (Gerngross & Puchta 2003). Furthermore, the distributional analysis with the test of the *null hypothesis*⁸ shows that this

8. The test of the null hypothesis serves to exclude other structural possibilities in order to determine whether the structure under investigation does indeed only occur in an invariant form in the speech sample.

form occurs invariantly in the learner's speech, i.e. it occurs without lexical or morphological variation (see Table 4).

Table 4. Example distributional analysis - learner C01 (Lenzing 2013: 166)

What do you X?	1
- What do you V X (What do you like for breakfast?)	1
Null Hypothesis:	-
What do he/she/it X?	0
What do we X?	0
What do they X?	0
What Ø you X?	0
What Ø he/she/it X?	0
What Ø we X?	0
What Ø they X?	0
What do Ø X?	0

On this basis, this question form is classified as a formulaic sequence and I argue that it is memorised as an invariant chunk by the learner. As pointed out above, I hypothesise that as far as formulaic sequences are concerned, no complete a-structure has been developed at this point.

Non target-like a-structures account for 50% of all question forms. These question forms deviate from the target-like a-structure in several ways: the learners produce questions with too many arguments, arguments that differ from the meaning the learner intends to express as well as question forms with an unclear a-structure. The different forms of deviations from the target-like a-structure are illustrated by the following examples.

(7) C08 She likes you spinach?

The question form in Example (7) contains an extra argument. As regards its underlying a-structure, there are two possibilities. Firstly, it could be the case that the a-structure comprises three arguments and looks as follows: *like ⟨experiencer, patient/theme, patient/theme⟩. However, as it can be inferred from the context that the question the learner intended to ask was 'Do you like spinach?'. This possibility seems to be highly unlikely. Hence, a more plausible explanation is adopted here: I assume that the expression 'she likes' constitutes merely a chunk that is stored as an unanalysed unit by the learner. The two arguments the learner intended to express, i.e. experiencer and patient/theme, are simply attached to this chunk. In line with this, I hypothesise that the arguments are directly mapped onto surface structure

without the assignment of grammatical functions (see Figure 9) and it is claimed that in this case, the a-structure has not been annotated for syntactic features yet.

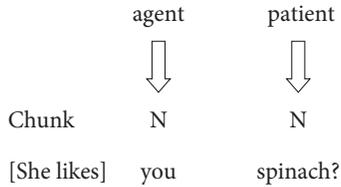


Figure 9. Direct mapping of arguments onto surface form (Lenzing 2013:216)

In the question form in Example (8), the arguments that are expressed are not the ones intended by the learner. Therefore, this question is classified as deviating in its a-structure from the target-like pattern.

(8) C10 I like spaghetti?

It could be argued here that the question form produced by learner C10 consists of a predicate with its corresponding arguments and thus, it could be classified as having a target-like a-structure. However, the detailed analysis reveals that the first argument is not the one the learner intended to express: it becomes clear from the context that the learner intended to ask the question: ‘Do **you** like spaghetti?’. For this reason, the utterance has been classified as deviating from the correct pattern.

In some cases, the a-structure cannot be determined which is due to the incomplete annotation of the lexical entry, as exemplified in (9):

(9) C24 What’s your eating?

According to the distribution of ‘eat’ in the question, it seems as if the verb ‘eat’ is used as a noun by learner C24. Hence, it is argued here that the lexical entry of the verb is not yet fully annotated, i.e. it is not annotated for its lexical category ‘verb’. As can be inferred from the context, the learner intended to ask ‘What do you like to eat?’.

7.3 Results statements grade 3

At the end of grade 3, the learners produce seven different types of lexical verbs and their speech exhibits a total of 11 verb tokens in statements. It is noteworthy that there is only one instance of a verb in the utterances that takes one argument; all other verbs take two arguments. Similar to the verbs occurring in question forms, none of the verbs in the statements had been translated. Thematically, the arguments are restricted to the four different roles agent, experiencer, patient/theme and locative.

Table 5. Lexical verbs in statements – types & token (Lenzing 2013: 218)

Verb types	Verb token	Number of arguments	Arguments – thematic roles
6	10	2	agent, experiencer, patient/theme, locative
1	1	1	agent

Table 5 presents the results of the detailed distributional analysis of a-structures occurring in the statements produced by learners at the end of grade 3. Table 6 shows that statements with lexical verbs are produced by seven learners. Out of these, only four learners produce statements with a target-like a-structure.

Table 6. Distributional analysis of a-structures in statements – grade 3 learners (Lenzing 2013: 218)

	C02	C03	C04	C08	C13	C17	C23
Formulae occurring in textbooks				1	1	1	
Non target-like a-structure:							
Argument in German		1					
Argument missing				1			1
Target-like a-structure	1		2	1		1	

About a quarter of the statements with lexical verbs in the speech samples (27%) are considered to be formulaic. Another 27% are classified as deviating from the target-like pattern and the remaining 46% of the statements display a target-like a-structure (see Lenzing 2013: 219).

The following example illustrates the occurrence of formulaic sequences with lexical verbs in the learners' speech sample:

(10) C08 I like spaghetti.

This utterance is classified as a formula as it can be unequivocally assigned to *Unit 4* in the textbook *Playway 3*. Additionally, the distributional analysis reveals its invariant occurrence in the learner data.

Statements with a non-target-like a-structure account for 27% of all statements with lexical verbs. These deviations consist of arguments that are expressed in German (33%) as well as missing arguments (67%) as is shown in the examples below (taken from Lenzing 2013: 220).

In Example (11), learner C03 expresses the second argument in German. In this case, it is the argument taking the thematic role of locative. I hypothesise that the a-structure is – at least partly – still annotated in the learner's L1:

- (11) C03 I live in Deutschland*
 *Deutschland = Germany
 a-structure: live ⟨agent, locative (L1)⟩.

Missing arguments account for two thirds (67%) of the deviations in a-structure. This means that not all arguments that are required by the predicate are expressed by the learner (see Example 12).

- (12) C23 Live in Paderborn.

As far as the underlying a-structure is concerned, there are two possibilities. It could be the case that the agent is missing in the a-structure which results in the hypothetical a-structure *live ⟨Ø, locative⟩. A second, more plausible possibility is that the agent is implicitly present in the semantic side of a-structure. The assumption here is that the agent is not expressed due to the incomplete annotation of syntactic features in the learner's a-structure.

Finally, target-like a-structures account for 46% of all statements with lexical verbs:

- (13) C11 I play the flute.
 a-structure: play ⟨agent, patient/theme⟩
- (14) C02 I like rolls {mit}*jam.
 *mit=with
 a-structure: like ⟨experience, patient/theme⟩

It is noteworthy that the preposition in Example (14) is expressed in German. However, as the constituent '{mit} jam' is an adjunct, it does not affect the underlying a-structure: like ⟨experiencer patient/theme⟩.

To summarise, the analysis of a-structures in question forms and statements in learners' speech after one year of instruction confirms my initial hypotheses outlined above. In particular, it shows that (1) both formulaic sequences and utterances with a non-target-like a-structure occur in the learners' speech samples and (2) there are differences between the a-structure of verbs in question forms and statements.

As far as differences between questions and statements are concerned, it can be observed that the learners produce a far greater amount of formulaic utterances in question forms (50%) than in statements (27%). Moreover, whereas 46% of all statements are classified as having a target-like a-structure, question forms with a target-like argument structure do not occur at all in the learner data. These findings support the hypothesis that the learners' a-structure is initially highly constrained. In particular, the results indicate that at the beginning of the L2 acquisition process, the a-structures are not completely annotated for the L2 in the learners' mental grammar and that this incomplete annotation applies especially

to the syntactic side of a-structure. In line with this, I hypothesise that at this early stage of L2 acquisition the arguments are not mapped onto grammatical functions. The inability to map arguments onto grammatical functions in f-structure is due to the constraints on the syntactic side of a-structure. Thus, the learners are able to merely map the arguments directly onto surface structure. This mapping process results in entirely linear surface structures. The results of the analysis show that it is exactly these linear surface structures that are present in the interlanguage of early L2 learners.

In keeping with this, some learners are able to produce statements with a target-like a-structure, as the production of structures adhering to canonical word order does not require the presence of grammatical functions. However, the situation is different for the generation of question forms. As predicted by the TOPIC Hypothesis, in order to process question forms, the learner has to acquire the non-linear mapping operations from c-structure to f-structure. In order to perform this kind of operation, it is essential to be able to assign grammatical functions to the respective arguments in a-structure. However, this assignment can only take place when the syntactic side of a-structure contains the relevant defining syntactic features which are a necessary prerequisite for a- to f-structure mapping. If the relevant a-structure is not annotated for its associated syntactic features, and subsequently, the arguments are not mapped onto grammatical functions, the non-linear mapping process from c-structure to f-structure cannot take place. The findings imply that the learners rely on direct mapping operations from arguments to surface form without making recourse to the grammatical functions at f-structure level.

7.4 Results grade 4 – developmental stages

The developmental stages of the learners after two years of instruction are summarised in Table 7.

Table 7. Overview of developmental stages of grade 4 learners (Lenzing 2013: 204)

Stage	Group 1						Group 2					
	C01	C02	C03	C04	C05	C06	C07	C08	C09	C10	C11	C12
6	-	-	-	-	-	-	-	-	-	-	-	-
5	-	-	-	-	-	-	-	-	-	-	-	-
4	-	-	-	-	-	-	-	-	+	-	-	-
3	-	-	-	+	-	-	+	+	(+)	-	(+)	(+)
2	(+)	+	+	+	+	+	+	+	+	+	+	+
1	+	+	+	+	+	+	+	+	+	+	+	+

(Continued)

Table 7. (Continued)

Stage	Group 3						Group 4					
	C13	C14	C15	C16	C17	C18.2	C19	C20	C21.2	C22	C23	C24
6	-	-	-	-	-	-	-	-	-	-	-	-
5	-	-	-	-	-	-	-	-	-	-	-	-
4	-	-	-	-	-	-	-	-	-	-	-	-
3	(+)	+	+	-	-	-	-	-	(+)	-	(+)	(+)
2	+	+	(+)	(+)	-	-	+	+	+	(+)	+	+
1	+	+	+	+	+	+	+	+	+	+	+	+

After two years of instruction, most learners have progressed in their development: only two of the learners, C17 and C18.2, are at stage 1 of their L2 development. Seven learners can be assigned to stage 2 of the PT hierarchy, as they produce 'SVO'-structures. Another three learners have acquired stage 3; they produce a sufficient number of 'stage 3' question forms. In the case of seven learners, it cannot be clearly determined whether they have acquired stage 3 of the processing hierarchy because their speech samples either exhibit only a small number of the respective features or it cannot be unambiguously determined whether the structures the learners produce are used productively or constitute formulaic patterns. Finally, one of the learners (C09) also produces 'stage 4'-structures.⁹

7.5 Questions grade 4

At the end of grade 4, the learners produce a larger amount of lexical verbs in questions than at the end of grade 3. The learner data contain eleven verb types and 49 verb tokens in question forms (see Table 8). However, five types had been previously translated from the learners' L1 by the interviewer. Similar to the results of grade 3, the thematic roles in question forms are restricted to agent, experiencer and patient/theme. An interesting finding here is that the learners also produce verbs that take only one argument.

Table 8. Lexical verbs in questions – types & token (Lenzing 2013: 223)

Verb types ¹⁰	Verb tokens	Number of arguments	Arguments – thematic roles
6 (4)	44 (8)	2	agent, experiencer, patient/theme
5 (1)	5 (1)	1	agent

9. For a more detailed discussion of the constraints at c-structure level, see Lenzing (2013).

10. The figures in brackets refer to the verb types and the verb tokens respectively that had been previously translated.

The detailed distributional analysis of the a-structure of the lexical verbs occurring in question produced by learners at the end of grade 4 is summarised in Table 9. It can be seen that question forms with lexical verbs occur in the data of 19 learners. Thirteen learners produce question forms with a target-like a-structure, six learners rely on previously-translated verbs in the production of question forms, a further five learners produce questions with ill-formed a-structures and seven learners make use of formulaic questions.

Table 9. Distributional analysis of a-structures in questions – grade 4 learners (Lenzing 2013:224)

	C03	C05	C06	C07	C08	C09	C10	C11	C12	C13
Formulae	1					1	1			
Translation	1									2
Non target-like a-structure:										
Argument in German					1				1	
Argument missing				1					1	
Too many arguments										
Predicate not intended one							1			
Target-like a-structure	1	1	1		1	1		6	4	2

	C14	C15	C16	C17	C18.2	C19	C21.2	C23	C24
Formulae				2	1		1	2	
Translation	2				1	2			1
Non target-like a-structure:									
Argument in German	1								
Argument missing									
Too many arguments			1						
Predicate not intended one									
Target-like a-structure	1	1					1	2	2

The results of the distributional analysis show that 49% of the questions display a target-like argument structure. Formulaic sequences account for 19% and in 14% of the question forms, the a-structure is not target-like. Interestingly, 18% of question forms with lexical verbs contain a previously translated verb (see Lenzing 2013:224).

Translated verbs are a new phenomenon that does not occur in the speech sample of the grade 3 learners. In this case, the learner asked the interviewer for

a translation equivalent of the verb before actually producing the utterance, as in Examples (15) and (16):

- (15) C18.1 Ski the mouse?
Hypothesised a-structure: ski ⟨agent⟩ (L1)
- (16) C14 Climb Max {Baum den Baum hoch}?
Hypothesised a-structure: climb ⟨agent, patient/theme⟩ (L1)
(Lenzing 2013:225)

I hypothesise that the underlying a-structure of question forms that contain previously translated verbs is not fully annotated in the learner's L2. Instead, I argue that the translated verb is only used in its L2 phonological surface form which implies that it still carries the annotations of the L1 equivalent so that the learner relies on the L1 a-structure.

The fact that the second argument in Example (16) is expressed in the learner's L1 (German) constitutes further evidence for the hypothesis that the a-structure of translated verbs is still annotated in the learner's first language.

As for the deviations from a target-like a-structure in question forms, the analysis shows that these are similar to those that were found at the end of grade 3. This means that the learner data contain instances of arguments that are expressed in the learner's mother tongue, of utterances with missing arguments or too many arguments as well as arguments that are expressed although they are not intended by the speaker.

The question forms that display a target-like a-structure are restricted both in terms of their lexical variation and their syntactic structure.

Table 10. Lexical restrictions in question forms with target-like a-structure – grade 4 learners (Lenzing 2013:226)

	C03	C05	C06	C08	C09	C11	
Lexical Verb	1	1	1	1	1	6	
	like	play	play	do you like X	like	4x do you like X 1x do you play X 1x swim	
	C12	C13	C14	C15	C21.2	C23	C24
Lexical Verb	4	2	1	1	1	2	2
	2x do you like X 2x do you play X	like, fly	play	like	do you like X	do you like X shine	like, walk

Half of the question forms with a target-like a-structure (50%) in the learner data consist of the two forms ‘Do you like X?’ and ‘Do you play X?’ (Lenzing 2013: 226) (See (17)):

- (17) C11 Do you like cake?
Do you play basketball?

The other half of the forms with a target-like a-structure is classified as fully productive, such as in Examples (18) and (19):

- (18) C14 Playing mouse and elephant football?
(19) C23 Is shining a sun?

Although these question forms are not target-like in terms of their syntactic structure, they display a target-like a-structure, as all the required arguments are expressed.

7.6 Results grade 4 statements

At the end of grade 4, an increasing number of statements with lexical verbs occur in the learner data. The verb types in statements amount to a total of 19 whereas the number of verb tokens is much higher, namely 131. It can be seen from Table 11 that the learners use the strategy of translation, i.e. they ask the interviewer for a translation of a German verb when they do not know the L2 equivalent. The thematic roles are mainly restricted to the roles of agent, experiencer and patient/theme. The thematic role of locative appears only twice in the speech sample with the verb *sit*, which was translated in both cases.

Table 11. Lexical verbs in statements – types & token (Lenzing 2013: 227)

Verb types	Verb tokens	Number of arguments	Arguments – thematic roles
8 (5)	85 (9)	2	agent, experiencer, patient/theme, (locative: translated verb)
11 (8)	46 (9)	1	agent

The distributional analysis in Table 12 shows that at the end of grade 4, 22 of the 24 learners produce statements with an underlying target-like a-structure. As for the remaining utterances, previously translated verbs occur in the speech samples of 17 learners and 13 learners also produce statements that deviate in their a-structure from the target-like pattern.

Table 12. Distributional analysis of a-structures in statements – grade 4 learners (Lenzing 2013:228)

	C01	C02	C03	C04	C05	C06	C07	C08	C09	C10	C11	C12
Formulae												
Translation		1	1	3	7	4		2		1		1
Non-target-like a-structure:												
Argument in German		1										
Argument missing						1			2	3	1	
Too many arguments										1		
Predicate not the intended one												1
Target-like a-structure		1	5	2	2	6	3	7	2	1	4	3
<hr/>												
	C13	C14	C15	C16	C17	C18.2	C19	C20	C21.2	C22	C23	C24
Formulae												
Translation	3	2	1	1		1			2	5	1	2
Non-target-like a-structure:												
Argument in German					2		1		2	1		
Argument missing			1	1			2					
Too many arguments												1
Predicate not the intended one												
Target-like a-structure	3	2	1	3		1	7	2	5	3	7	3

To summarise the results from Table 11, the majority of statements the learners produce (55%) include a target-like a-structure. In 16% of all utterances, the a-structure is not target-like and in 29% of all cases, the verb had previously been translated.

The detailed analysis shows that similar to the results for the question forms discussed above, the deviations from a target-like a-structure in statements consist of missing arguments, arguments that are expressed in the learner's L1 (German), too many arguments or arguments that are expressed although their meaning is not intended by the speaker.

In sum, the results of the analysis of grade 4 learners show that whereas the learners still use formulaic sequences in question forms (19%), they do not produce any formulaic structures in statements anymore. Furthermore, the analysis reveals that whereas deviations in a-structure in question forms and statements occur at a similar rate (14% vs. 16%), the percentage of utterances with a target-like a-structure is higher in statements (55%) than in question forms (49%). Again, these findings support the initial hypotheses of this study, as they indicate that the grammatical system of the L2 learners is still constrained at the level of a-structure after two years of formal instruction.

In the following, the results of the two analyses of learners' speech are compared with each other.

8. Comparison results grade 3 – grade 4

In terms of the developmental stages the learners have reached after one and after two years of instruction, it can be seen that the learners' development in their second language is in line with the predictions made by *Processability Theory*. Nearly all learners progressed and no stages were skipped in this process. Whereas the majority of the learners were located at stage 1 at the end of grade 3 and their speech was characterised by formulaic sequences, single words and idiosyncratic utterances, they reached stage 2 or 3 after two years of instruction in English, i.e. they were able to produce sentences with canonical word order (i.e. SVO) as well as a limited range of question forms.

The analysis of a-structure in both question forms and statements of learners after one and after two years of instruction is summarised in Table 12.

Table 13. Comparison a-structures grade 3 – grade 4 (Lenzing 2013:233)

A-structure	Grade 3		Grade 4	
	Questions	Statements	Questions	Statements
Formulaic structures	50%	27%	19%	–
Deviations in a-structure	50%	27%	14%	16%
Translation	–	–	18%	29%
Complete a-structure	–	46%	49% (partly restricted)	55%

After two years of instruction the a-structures in learner utterances differ in several respects from the ones after one year of instruction. These differences apply to both question forms and statements.

As for statements, it becomes clear that whereas formulaic utterances account for 27% of the statements the learners produce at the end of grade 3, these do not occur at all in the learners' speech at the end of grade 4. This indicates a clear development away from formulaic sequences towards a more productive use of statements. Furthermore, the number of statements with a fully productive target-like a-structure increases slightly. Whereas at the end of grade 3, half of the statements (46%) are considered to have a target-like a-structure, this applies to 55% one year later.

The comparison of question forms shows that the formulaic structures decrease in the course of second language development: whereas at the end of grade 3, 50% of all questions are considered to be formulaic, this applies to only 19% after two years of instruction. Moreover, the learners start to produce question forms with a complete a-structure at the end of grade 4. Again, this indicates a move away from formulaic speech towards a more productive use of the second language.

9. Conclusion

The results of the present study indicate that the initial mental grammatical system of early L2 learners is highly constrained, and that these constraints apply to the level of constituent structure as well as to level of argument structure, which was the focus of the analysis presented in this paper.

The overview of the developmental stages of the learners shows that at the level of c-structure, the learners' speech production after one year of instruction is mainly restricted to single words, formulaic sequences and idiosyncratic utterances. Although the longitudinal study reveals a clear developmental progress in this respect, the results of the analysis of grade 4 learners indicate that their grammatical system is still restricted after two years of instruction.

The analysis of a-structure showed the following two main results: (1) the early learner utterances contain both formulaic sequences and non target-like a-structures and (2) there are differences between the a-structure of question forms and the one of statements. This means that the learner data display more formulaic sequences and non target-like a-structures in questions than in statements. These findings lend strong support to the hypothesis that initially, a-structure is not fully developed in the L2 acquisition process. In particular, the differences in a-structure between question forms and statements support the claim that the restrictions in a-structure especially apply to the syntactic features that are a necessary prerequisite for the mapping process of a- to f-structure. In line with this, the study presented

here provides evidence for the hypothesis that the learners rely on direct mapping processes from arguments to surface form. As pointed out above, these direct mapping operations can be performed without direct access to the grammatical functions encoded in f-structure, as put forward in the MCH.

All in all, the results of the longitudinal study indicate a clear development towards less formulaic speech and more complete a-structures in the course of L2 acquisition, since the postulated constraints are gradually relaxed as learning progresses. This development is in accordance with the predictions made by PT as well as the MCH.

References

- Alsina, A. (1996). *The role of argument structure in grammar: Evidence from Romance*. Stanford, CA: CSLI.
- Bresnan, J. (2001). *Lexical-Functional Syntax*. Malden, MA: Blackwell.
- Bresnan, J., & Kanerva, J. (1989). Locative inversion in Chichewa: A case study of factorization in Grammar. *Linguistic Inquiry* 20(1): 1–50.
- Dalrymple, M. (2001). *Syntax and semantics. Lexical functional grammar. Vol. 34*. San Diego, CA: Academic Press. doi:10.1163/9781849500104
- Falk, Y. (2001). *Lexical-functional grammar: An introduction to parallel constraint-based syntax*. Stanford, CA: CSLI.
- Fabri, R. (2008). Lexical functional grammar. In J.-U. Keßler (Ed.), *Processability approaches to second language acquisition: Development and second language learning* (pp. 31–67). Newcastle upon Tyne: Cambridge Scholars.
- Gass, S., & Mackey, A. (2007). *Data elicitation for second and foreign language research*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Gerngross, G., & Puchta, H. (2003a). *Playway 3 rainbow edition. Teacher's book*. Innsbruck: Helbling/ Leipzig: Klett.
- Gerngross, G., & Puchta, H. (2003b). *Playway 4 rainbow edition. Teacher's book*. Innsbruck: Helbling & Leipzig: Klett.
- Johnstone, R. (2000). Context-sensitive assessment of modern languages in primary (elementary) and secondary education: Scotland and the European experience. *Language Testing* 17(2), 123–143.
- Lenzing, A. (2013). *The development of the grammatical system in early second language acquisition – The Multiple Constraints Hypothesis*. Amsterdam: John Benjamins. doi:10.1075/palart.3
- Mackey, A., & Gass, S. (2005). *Second language research. Methodology and design*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Pienemann, M. (1998). *Language processing and second language development. Processability Theory*. Amsterdam: John Benjamins. doi:10.1075/sibil.15
- Pienemann, M., & Mackey, A. (1993). An empirical study of children's ESL development and Rapid Profile. In P. McKay (Ed.), *ESL development. Language and literacy in schools* (Vol. 2; pp. 115–259). Canberra: Commonwealth of Australia and National Languages and Literacy Institute of Australia.

- Pienemann M. & Lenzing, A. (2015). Processability Theory. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition. An introduction*, 2nd edition (pp. 159–179). Mahwah, NJ: Lawrence Erlbaum Associates.
- Pienemann, M., Di Biase, B., & Kawaguchi, S. (2005). Extending Processability Theory. In M. Pienemann (Ed.), *Cross-linguistic aspects of Processability Theory* (pp. 199–251). Amsterdam: John Benjamins. doi:10.1075/sibil.30.09pie
- Roos, J. (2007). *Spracherwerb und Sprachproduktion: Lernziele und Lernergebnisse im Englischunterricht der Grundschule*. Tübingen: Gunter Narr.

Question constructions, argument mapping, and vocabulary development in English L2 by Japanese speakers

A cross-sectional study

Satomi Kawaguchi

School of Humanities & Communication Arts, Bilingual Research Lab & MARCS Institute, Western Sydney University

This study investigates the relationship between vocabulary size (Nation and Beglar 2007) and syntactic learning in English as a second language (ESL) using the framework of Processability Theory (PT, Pienemann 1998, Pienemann et al. 2005). In particular, the study focuses on the syntactic development of question sentences and argument mapping in conjunction with the learner's current vocabulary size. Nine adult Japanese L1-English L2 speakers in Australia were selected out of a total sample of 22 who sat for the vocabulary size test, three each from Top, Middle and Low vocabulary sizes, to perform two language production tasks: (1) a 'spot the differences' task, used for speech profiling and (2) a translation task involving a range of verb categories including unaccusative verbs, psych verbs, as well as passive and causative constructions. The linguistic production of each informant was analysed against PT syntactic stages (Bettoni & Di Biase, 2015) in question sentences and argument mapping. Results suggest that vocabulary and syntactic development progress hand-in-hand. However Low and Mid vocabulary size ESL learners have problems in specific areas of syntax. High vocabulary learners, on the other hand, were able to cope with the whole range of verbs and syntactic constructions investigated in this study. Question sentences and argument mapping were found to be key indicators of ESL learners' syntactic development. The broad goal of this investigation is to promote intermediate-advanced learners in ESL.

1. Introduction

While this study focuses on the relationship between ESL speaker's vocabulary size and syntactic ability within the framework of Processability Theory (PT, Pienemann 1998) the broad goal of my investigation is to promote further

development of intermediate-advanced L2 learners in Australia. A key issue in learning second languages (L2) is that most learners do not progress beyond the intermediate level. Overseas students in Australia have significant difficulties with their English over the course of their endeavour (Zhang & Mi 2010) and even after completion of their university studies (Birrel 2006). The low rate of success observed beyond beginner-intermediate levels in L2 performance is a well-known but little understood problem: how can L2 learners progress beyond intermediate levels of performance? This proposition holds also with migrants in Australia. Immigration Minister Scott Morrison stated that migrants should take more than one English proficiency test if they want to stay in Australia longer-term and that migrants' English proficiency should not remain static but should develop and improve over time if they aim to live and work in Australia (Australian Associated Press, 17 July 2013). But how can the learner 'develop and improve' their English L2 to a high proficiency level?

In the field of second language teaching, while the L2 beginners tend to receive more support from the teacher to learn the language, more advanced students are expected to be more independent and resourceful for their L2 learning (Harmer 2007). The second language acquisition (SLA) field is aware of the difficulties for learners to achieve high levels of effective communication in their L2, but so far it has directed the bulk of its research effort towards understanding the beginning and intermediate stages of acquisition, tacitly assuming, perhaps, that once the early obstacles along the path are overcome the second language learner may well be in a position to further their L2 knowledge and skills in a more self-sufficient manner. To begin unravelling this question we need to identify which aspects of language knowledge and skills appear to be consistently difficult for intermediate learners, whether these difficulties occur in similar areas across languages, at what point in L2 development, and whether training focusing on these areas is beneficial to learners in a measurable way. Several studies of advanced L2 users point to one persistent area of difficulty: the integration of syntactic and pragmatic information (Hopp 2007; Sorace 2003), which occurs when speakers do not rely exclusively on default syntactic and lexical options.

The current study focuses on the development of: (1) question sentence constructions and (2) argument mapping including specific verb lexical category. These two aspects of English development are selected because they play central roles in syntactic construction and argument realization (c.f., Levin & Rappaport Hovav 2005). For instance, word order is specific in English depending on whether the sentence is either declarative or interrogative. Also, English requires a variety of syntactic frames to construct both Yes/No and Wh-questions such as the selection of auxiliary verbs which carry SUBJ information and tense information

(e.g., do/does/did). Further, the verb specifies the sentence argument structure. For example, the verb *kick* requires two participants in the event, i.e., an *agent* and a *patient*, but the verb *cry* requires *agent* as the sole argument; the verb *be boring* takes *theme* as a subject (as in *his lectures are boring*) while the verb *be bored* takes *experiencer* as its argument (as in *I am bored*). These two items are often confused by Asian ESL learners (Kawaguchi 2013), who may come up with *I am boring* (*by his lectures*). The way in which the verb organizes its arguments may differ from language to language, and since Japanese and English contrast typologically in significant ways Japanese L1-English L2 offer an ideal hypothesis-testing linguistic constellation. This cross-sectional study of adult, Japanese L1-English L2 looks at, in particular, the relationship between learner's vocabulary size in English, measured according to the instrument developed in Nation and Beglar (2007), and syntactic development, measured by English L2 developmental stages as defined by PT. Thus the research questions to be answered through the study are:

1. Is there a relationship between the learner's vocabulary size and his/her syntactic ability to produce question sentences in English L2?
2. Is there a relationship between the learner's vocabulary size and their grammatical ability to produce different types of syntactic frames for a range of verb types?

2. Vocabulary size and language acquisition

Previous studies on L2 lexicon and lexical acquisition (e.g., Nation 2001; Laufer & Hulstijn 2001; Kroll & Tokowicz 2001) offer insights into lexical acquisition in such areas as second language acquisition and the language professions, e.g., translation and interpreting. However, a key issue with these studies is that they tend to treat all vocabulary items in a statistically uniform way. Yet, many modern theories of grammar (Bresnan 2001; Culicover & Jackendoff 2005; Van Valin & la Polla 1997; Van Valin 2005) assume that syntax is driven by the lexicon. Some researchers believe lexical size is one way of indicating L2 learner's proficiency level especially in reading and listening (Nation 2001; Mochida & Harrington 2006). It can also be used as a reference point for phonological, morphological and syntactic development, much as mean length of utterance (MLU) is used in first language acquisition. However, what lexical size seems to measure is the learner's semantic knowledge of a word but not necessarily its grammatical and combinatorial features and their values (i.e., the full *lemma* in Levelt 1989;

Bock & Levelt 1994; and Levelt, Roelofs, & Meyer 1999's terms), and therefore its actual use in connected oral and written production. In fact, many current cognitive approaches to SLA (e.g., DeKeyser 2007; Pienemann 1998; VanPatten & Houston 1998) show that grammatical knowledge (i.e., declarative knowledge) is different from language procedural skills for speech production and comprehension. Then learner's lexical size based on word frequency lists may not be sufficient to predict learners' *productive ability* in L2 syntax. This study may be able to contribute to making lexical size instruments better connected to overall L2 development, and hence point towards additional instruments to resolve the limitations of the lexical size test.

3. Processability Theory (PT) and its hypotheses

The framework of this study, Processability Theory (PT), is a universal theory of second language acquisition based on general human cognition such as speech processing architecture, lexical access and memory capacity. The theory explains why the second language learner develops his/her language in a specific order. Because of its psychological and typological plausibility, PT is a valuable instrument for describing, predicting, and accounting for the development of L2 syntax and morphology in speech of typologically different second languages.¹ For the processing perspective which also accounts for a set of key psychological aspects, PT is based on Levelt's (1989) speech generation model, which shares many basic notions with Kempen and Hoenkamp's (1987) Incremental Procedural Grammar (IPG), which is a performance production grammar, and Lexical Functional Grammar (LFG) (Kaplan & Bresnan 1982; Bresnan, 2001 among others) as a psychologically and typologically plausible formal grammar. All three theories (i.e., Levelt's model, IPG and LFG) agree on the assumption that grammar is lexically driven. Based on this assumption, i.e., that grammar is lexically driven, as well as on the incremental nature of speech processing, Kempen and Hoenkamp (1987) proposed that grammatical encoding is activated in the following order in the formulator:

1. For example, Pienemann, *Language processing* (English); Sakai, 'An analysis of Japanese'; Di Biase and Kawaguchi, 'Exploring typological plausibility' (Italian); Pienemann and Håkansson, 'A unified approach' (Swedish); Mansouri, 'From emergence to acquisition', and 'Agreement morphology' (Arabic); Zhang, 'A processability approach', and 'Processing constraints' (Chinese); Iwasaki 'The acquisition of Japanese'; Kawaguchi 'Argument structure', and 'Lexical mapping theory' (Japanese).

- (1) 1. lemma,
2. the category procedure (lexical category of the lemma),
3. the phrasal procedure (instigated by the category of the head),
4. the S-procedure and the target language word order rules,
5. the subordinate clause procedure – if applicable.

Pienemann (1998) hypothesised that there is a hierarchical relationship for the acquisition of the processing resources by the L2 learner which follows the same sequence as the activation in the production process listed in (1) above. This is because the processing resources to be acquired form an implicational hierarchy in the encoding process. Based on Levelt (1989), Pienemann (1998) claims that the processing resources at the lower level are a prerequisite for the higher level ones. The acquisition of these resources allows for staged developmental sequences in L2 syntax. Based on LFG's lexical mapping theory (Bresnan 2001) PT further postulated two hypotheses explaining interlanguage development: the Topic Hypothesis and the Lexical Mapping Hypothesis (Pienemann, Di Biase, & Kawaguchi 2005). These two hypotheses are relevant to this study as the Topic Hypothesis, currently subsumed under the Prominence hypothesis to encompass also Focus (Bettoni & Di Biase 2015) accounts for the development of questions sentences while the Lexical Mapping Hypothesis concerns the development of argument mapping between thematic roles (e.g., Agent, Patient) and grammatical functions (e.g., Subject, Object) in sentence constructions. In this study, I utilise Bettoni & Di Biase's (2015) English L2 stages of question sentences. The differences between English L2 stages in the original PT (Pienemann 1998) and Bettoni & Di Biase (2015) in the analysis of English L2 question sentences are a consequence of current development of LFG. While original PT (Pienemann 1998) uses Kaplan and Bresnan (1982) focusing on word order in *c*-structure and constraint equations, Bettoni and Di Biase (2015) as well as Di Biase et al. (2015) use Bresnan (2001) and Dalrymple (2001) for formalisation which incorporates information structure and grammaticised discourse functions TOP and FOC. Note that incorporation of information structure to explain English L2 stages does not invalidate the original explanations provided by Pienemann (1998). Instead, adding the notion of TOP and FOC rather explores a new dimension in explaining syntactic phenomena. Another difference between Pienemann (1998) & Pienemann et al. (2005) and Bettoni and Di Biase (2015) may be the separation of declaratives from questions (and yes/no from content questions) in testing the stages hypotheses. Although Bettoni and Di Biase (2015) and Pienemann (1998) use somewhat different interpretations and labelling for the syntactic stages, the actual developmental sequences in English appear to be the same.

4. The discourse functions hypothesis and development of Y/N and Wh-questions

The Prominence Hypothesis states that:

In second language acquisition learners will initially not differentiate between grammatical functions (GFs) and discourse functions (DFs), for example, between SUBJ and TOP. Differentiation begins when an element such as an XP, or other lexical material, is added to the canonical string in a position of prominence in c-structure, that is, the first in the sentence. This element may be TOP in declaratives or FOC in interrogatives leaving, crucially, the canonical string unaltered. At the next stage, learners will be able to construct noncanonical strings assigning prominence to any constituent in an unequivocal way. (Bettoni & Di Biase 2015:63)

This hypothesis predicts Y/N questions and Wh-questions in English to be acquired following the order as schematically presented in Table 1 and Table 2 respectively (beginning from lower to higher rows in the tables). These stages are attested in a longitudinal study (Yamaguchi 2008, 2009, 2010). As can be seen in Table 1, the first stage of interrogative is lemma access realised as single words and formula (e.g., *Coffee?*) with QUE^P where the superscripted “p” indicates that the question modality is expressed only prosodically (with raising intonation). Exclusively prosodic marking [QUE^P] continues also at the second, i.e., Canonical Order, stage (*Tom is happy?*) where the question modality is still entrusted to the prosodic envelope of the Canonical Order expression, which would otherwise express a declarative. At the next stage some lexical material (beside the prosody) starts to distinguish questions. This lexical material is a kind of particle which is preposed (in topical position) hence outside of the Canonical Order block “QUE particle + Unmarked alignment” as in *do they have cat?*. The highest stage (in single clauses) involves marked alignment as in *have you tried pizza?* where English canonical order SUBJ-VERB-OBJ is superseded by the auxiliary occupying the particle position. The Auxiliary is different from the particle because it carries subject person marking or tense information. In short, the learner progresses from single words to canonical order (unmarked alignment) then to particle and finally to marked alignment.

As for the developmental stage of content questions (see Table 2), a similar progression occurs but this time the Focus is expressed by the Question word (Wh-). The first stage is also single words and formulaic expressions involving question words, such as *what?*, *how much?* or *how are you?* The next stage may be realised as In-Situ WH-question constructions where the canonical, unmarked alignment is preserved, such as *John eat what?* common in languages such as Chinese or Japanese. Not all learners go through this stage which may turn out to be ungrammatical (except in marked, specific discourse). The next stage is

QUE(stion) word + Unmarked alignment such as *what he eat?* which also exhibits ungrammatical construction in English. The highest stage (in single clause) in Wh-questions involves marked alignment as in *when are you going?*. English indirect question sentence such as *I wonder where the station is* involves subordination and inter-clausal processing (cf Pienemann & Kessler 2011), which I do not discuss in this paper.

Table 1. Developmental stages for English syntax: Y/N questions
(After Bettoni & Di Biase 2015)

Stage	Structure	Example
4. marked alignment	AUX _{QUE} SUBJ V (O) MOD _{QUE} SUBJ V (O) <i>have</i> _{QUE} SUBJ OBJ copula _{QUE} SUBJ predicate	<i>have you tried pizza?</i> <i>can Ann swim?</i> <i>have you a boyfriend?</i> <i>is Joan happy?</i> <i>are you there?</i>
3. QUE(stion) particle + unmarked alignment	QUE [canonical order]	<i>do they have cat?</i> <i>is your man have a red hat?</i> <i>is Mary is happy?</i>
2. unmarked alignment	[QUE ^P canonical order]	<i>dog eating the doughnut?</i> <i>you like pizza?</i> <i>you are there?</i> <i>Tom is happy?</i>
1. lemma access	[QUE ^P single words] [QUE ^P formulas]	<i>Jim happy?</i> <i>coffee? going?</i>

QUE^P feature is only prosodic

Table 2. Developmental stages for English syntax: Constituent questions
(After Bettoni & Di Biase 2015)

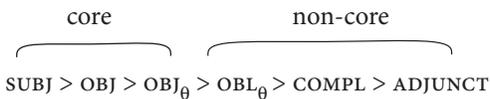
STAGE	STRUCTURE	EXAMPLE
4. XP _{FOC} MARKED ALIGNMENT	WH _{QUE} AUX SUBJ VO WH _{QUE} MOD SUBJ V (O) WH _{QUE} copula SUBJ	<i>what has Tom eaten?</i> <i>where did Joan go?</i> <i>when are you going?</i> <i>what can Mary do?</i> <i>where are they?</i> <i>what is this?</i>
3. XP _{FOC}	WH _{QUE} canonical order	<i>what he eat?</i> <i>when you go?</i> <i>where Joan is?</i>
2. UNMARKED ALIGNMENT	in-situ _{QUE} [Canonical Order]	<i>Joan eat what?</i>
1. LEMMA ACCESS	single word formula	<i>what? what colour?</i> <i>how much is it?</i>

5. The Lexical Mapping Hypothesis

Many linguists (e.g. Bresnan 2001; Foley & Van Valin 1984; Givón 1984; Jackendoff 1972) have suggested a universal hierarchy of thematic roles, as in (2). This hierarchy orders the relative prominence of the arguments of a predicator: the higher the level in the hierarchy, the more cognitively prominent the argument. Grammatical functions also have a hierarchical relationship according to their prominence, as in (3). All core functions are more prominent than non-core functions. So the most prominent role is universally the agent. This means that the agent is more likely to be encoded as a core-argument, such as subject, rather than non-core argument (Bresnan 2001). On the other hand, the locative role, located at the bottom of the hierarchy, is less prominent and likely to be encoded as a non-core-argument rather than core argument.

- (2) Thematic hierarchy (Bresnan 2001:307)
 Agent > Beneficiary > Experiencer/Goal > Instrument > Patient/Theme >
 Locative

- (3) Relational hierarchy (Keenan & Comrie 1977, referred in Bresnan 2001:96)



Based on these universal thematic and relational hierarchies, PT's Lexical Mapping Hypothesis states as follows:

Second language acquirers will initially map the highest available role in the thematic hierarchy (e.g., agent, experiencer) onto minimally specified SUBJ/TOP. We call this default mapping. Next, they learn to add further arguments mapped onto grammatical functions (GFs) differentiating them from SUBJ (and OBJ, if present). They may also learn some exceptional verbs at this second stage. Finally, they learn to impose their own perspective on events, that is, to direct the listener's attention to a particular thematic role lower in the hierarchy by promoting it to SUBJ, and defocus the highest role by mapping it onto a GF other than SUBJ, or suppress it altogether. At this last stage learners may add further roles information regarding causality, benefit, or adversity. They may also add to their lexicon particular subsets of Vs, such as unaccusatives, as well as further intrinsically exceptional Vs requiring their own mapping schema. This final stage we call nondefault mapping.
 (Bettoni & Di Biase 2015: 68)

The Lexical Mapping Hypothesis, then, predicts that the initial syntactic structure that learners construct as soon as they are able to produce utterances of more than one word will utilize canonical mapping. This contributes to the realisation of default mapping structures in the L2 which rely on the association Agent-Subject

and Patient-Object appearing in a fixed position.² Such association, in line with other acquisition theories (e.g. Pinker 1984; Slobin 1982), is assumed to require the least processing effort (e.g. Pienemann et al. 2005). From a psycholinguistic point of view, *agent* (rather than *patient* or *theme*) is the most prominent participant role in an event (Jackendoff 1972), and from a grammatical point of view, Subject is the most prominent grammatical function (Keenan & Comrie 1977). Thus the Agent-Subject association is the most harmonious because the most prominent participant role is mapped onto the most prominent grammatical function taking the most prominent (first) position (Choi 2001). This is schematically represented in Figure 1 with the English example *the cat ate the fish* where the Agent ⟨cat⟩ is mapped onto the grammatical function (GF) Subject while the patient ⟨fish⟩ is mapped on the Object GF. By contrast, in passive sentences, exemplified in Figure 2, the highest thematic role (i.e. Agent) is suppressed. However, the suppressed *agent* may appear as Adjunct. Hence passive constructions are clear cases of non-canonical mapping. L2 learners, however, find this difficult at first. The lexical mapping predicts that they will learn this alternative mapping only after canonical mapping is in place.

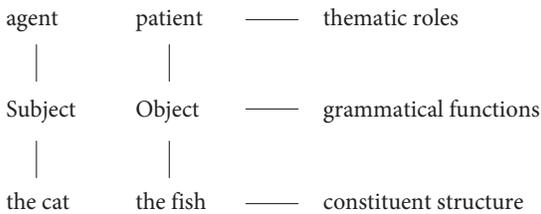


Figure 1. Active mapping: *the cat ate the fish*

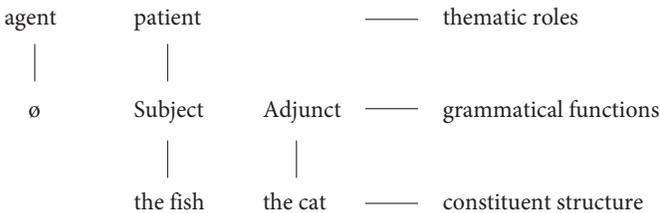


Figure 2. Passive mapping: *the fish was eaten by the cat be eaten ⟨x⟩*

2. Out of the six possible ways of ordering Subject, Object and Verb in languages, SVO, SOV and VSO “are overwhelmingly more frequent, reflecting the universal tendency for the subject to precede the Object” (Comrie, Matthews, & Polinsky, 2003). See also Greenberg (1966), Tomlin (1986).

Speakers use more costly non-canonical structures such as passive because they are linguistic devices to attribute prominence to thematic roles other than the Agent. For example, the passive is a linguistic alternative way to construct a verbal message to place prominence on Patient rather than Agent to facilitate comprehension (Levelt 1989) and allows the speaker to impart different perspectives on discourse world situations (Payne 2011). Next, I exemplify default and non-default mapping structures used in my study.

Canonical mapping: The sentence in (4) represents a typical canonical mapping construction with a transitive verb *break* which requires two arguments where the more prominent role, the Agent, is mapped on the Subject and the less prominent role, the Patient, is mapped on the Object grammatical function. Also some intransitive verbs³ (the unergative ones) whose sole argument is typically an Agent or Experiencer – a role high in the thematic hierarchy – map on the Subject.

- 4) canonical transitive
Break ⟨Agent, Patient⟩ *I broke the stick*

Non-canonical mapping: A typical case of non-canonical mapping is the passive construction, explained above. This type of non-canonical mapping is usually called ‘structural’ because the alternative lexical entry (*be eaten*, versus active *eat*) creates a structural frame which is regular and predictable. Causative constructions are similarly non-canonical mapping structures, which are also regular and predictable alternative constructions (see Pienemann et al. 2005, and Kawaguchi 2009 for Japanese causatives). Other non-canonical mappings are created ‘lexically’ in the sense that they are intrinsically required by the lexical verb, hence they are neither regular nor predictable so they need to be learned case by case. Characteristically, these verbs map hierarchically lower thematic roles, e.g., Theme, on the Subject. For instance, with the unaccusative alternative of the verb *open* in (5), the hierarchically lower role, Theme, is mapped on the Subject while the Agent role of the eventuality of ‘opening’ is actually excluded from the scene altogether. Another group of verbs which build non-canonical mappings in English are the so-called *Psych verbs* (c.f. White et al. 1998). For example, the verb *frighten*, in (6), requires the Theme *her screams* (i.e. a lower role in the thematic hierarchy) to be mapped

3. Intransitive verbs, which require only one argument, are divided into unergatives and unaccusatives (Burzio 1986). These classifications are based on the thematic role that the sole argument carries in the sentence. The argument of unergative verbs typically bears an agent or experiencer role as in (a) while that of unaccusative verbs typically bears a theme or patient role as in (b).

- (a) Tom cried (Unergative); (b) The window broke (Unaccusative)

on the Subject while *John*, the Experiencer (i.e. a higher role in the hierarchy), is mapped on the less prominent grammatical function Object.

- (5) Unaccusative verb⁴
 open ⟨Theme⟩ *The door opened suddenly*
- (6) Psych Verb: OBJ Experiencer (OE)
 frighten ⟨Theme, Experiencer⟩ *Her screams frightened John*

Hence non-default mapping can be generated either lexically, as exemplified in (5) and (6) or structurally as in passives and causatives discussed above. Table 3 summarises the developmental stages based on the Lexical Mapping Hypothesis. The intermediate stage between default and nondefault mapping (i.e., Default mapping + additional argument) is not discussed in this paper because the tasks used were not designed specifically to elicit ditransitives and oblique arguments.

Table 3. Developmental stages for English syntax based on the Lexical Mapping Hypothesis – Declaratives (after Pienemann, Di Biase, & Kawaguchi 2005: 246)

Stage	Structure	Example
nondefault mapping	exceptional verbs, passives, causatives, etc.	<i>Silvie pleases Jacques</i> <i>the blue fish is eaten by the green fish</i> <i>she let him sleep longer</i>
Default mapping + additional argument	Ditransitive Canonical sentence + OBL	<i>Tom give Mary a pen</i> <i>I showed the picture to my friends.</i>
default mapping	e.g., agent-event-patient; experiencer-event-theme & canonical word order	<i>the green fish eats the blue fish</i> <i>Jacques likes Silvie</i>
lemma access	single words; formulas	<i>station, here</i> <i>my name is Pim</i>

6. Study

This section describes the study design including informants, procedure, tasks and data analysis methods to answer the research questions presented earlier.

4. There are alternating (e.g., *close*, *break*) and unalternating accusative verbs (e.g., *arrive*, *appear*) in English (see Hirakawa 2003). The former involve non-canonical mapping while the latter build canonical mapping.

6.1 Informants

The informants in this study were 22 Japanese L1 speakers of English L2 (five male and 17 female) aged between 20 and 56 years (mean 31, SD 9.9) with lengths of stay in Australia ranging from 9 days to 27 years. They include: working holiday participants, university students (all undergraduate, MA and Ph.D.), business people and their wives and one professional translator. Adult informants of varying lengths of stay may provide a wide range of attainment in English L2. An 18-year-old simultaneous bilingual first language speaker of English and Japanese, born in Australia by Japanese native speaking parents, participated as a control since one of the tasks involved Japanese to English translation. In order to ensure the informants' anonymity, codes such as JA1, JA2 were assigned.

6.2 Procedure

The following procedure was implemented.

1. Conduct a vocabulary size test with 22 Japanese L1-English L2 speakers in Australia.
2. Analyse the vocabulary test results and choose three informants each from Top, Mid(dle) and Low (i.e., bottom) vocabulary size groups, nine in Total. These three groups of informants enable us to compare syntactic abilities of the English L2 speakers of different vocabulary sizes.⁵
3. Interview each the nine informants using a profiling task to check their syntactic developmental stage, particularly with question sentences, based on Processability Theory.
4. Conduct a translation production task involving a selection from different types of default and non-default mapping.
5. The data obtained through the profiling and translation tasks is then analysed against PT predictions. This involves full distributional analysis followed by implicational scaling for measuring language development thus providing the framework within which the relationship between vocabulary size and syntactic stage is examined.

5. Note this study looks at adult English L2 speakers in Australia who have completed compulsory English studies at least 6 years in Japan. Thus, High, Mid and Low may not correspond to general definitions of learners' lexical ability.

6.3 Tasks

a. Vocabulary size test

The Nation and Beglar (2007) vocabulary size test that measures vocabulary knowledge (for comprehension) up to the 14,000 word families level was used in order to identify three informants from each vocabulary size ranges (top, mid and bottom) among the 22 informants. This vocabulary size test is well supported (e.g. Nation 2006) since a significant correlation between vocabulary size and receptive language abilities (i.e. reading and listening) has been established. It is interesting to test whether productive (as opposed to receptive) language ability also has a relationship to vocabulary size.

b. Profiling task

A profiling task was used to elicit production data to analyse the participant's PT stages. First, a short interview about the participant in general was carried out. This aims to elicit the participant's bio data (such as ESL instruction, length of stay in Australia, etc.) but also aims to elicit various syntactic structures. The interview was followed by a 'spot the differences' task, which aimed to elicit various English question sentences, which served to identify the participants' ESL PT stage. In this task, the participant and the interviewer each took one of two fairly similar pictures each that differed in some (around 10) details. For example, both pictures may depict a public garden but one picture has one dog while the other has two. The task for the participant was to find differences between the two pictures by first describing their picture and then by asking questions about the interviewer's picture. Neither participant can see the other's picture. In this particular study the research assistant, an MA holder in Applied Japanese Linguistics who is a native speaker of Japanese with an advanced command in English, acted as interviewer.

c. Translation task

The third task was a written translation task eliciting syntactic production involving default and non-default mappings. There are not many studies of such productive abilities in the field of second language acquisition; with the exceptions of Hirakawa (2003) and White et. al. (1998) most tasks involve either comprehension tests or grammatical judgment tests. It is a challenge to elicit a range of speech production involving transitive and intransitive contrasts from L2 learners, which constitute an important part in testing the Lexical Mapping Hypothesis. PT is a SLA theory based on speech processing. Therefore, PT studies traditionally utilised online speech elicitation tasks. However, the Steadiness Hypothesis (Pienemann 1998) across modality has been addressed by a few studies in recent years. For example, in Rahkonen and Håkansson (2008)

L2 Swedish ‘writing’ was tested, the results of which showed that acquisitional patterns may differ significantly according to whether corpora include semiformal writing closer to speech or formal writing. Further, Håkansson and Norrby (2007) showed that both oral and written development follow the PT hierarchy. Kawaguchi (2015) also shows that L2 Japanese learners’ syntactic and morphological development in text chat (i.e., informal writing) follows PT hierarchy. These studies pave the way for the use of informal written production data to measure L2 acquisition in PT.

In this task, the informants were asked to translate 25 Japanese sentences into English and were instructed to use a particular English verb in their translation for each sentence (see Appendix A). Six of these 25 sentences, which involve a variety of constructions such as raising and subordination, were not used in the present study. The 19 verbs tested in the translation task (summarised in Table 4) contain five default and fourteen non-default structures (six lexical non-default and eight structural non-default structures).

Table 4. The 19 English verbs targeted in the translation production task

Canonical transitive (n=5)	Non-canonical			
	Lexically non-canonical		Structurally non-canonical	
	Intransitive (Unaccusative) (n=3)	Transitive (Psych Verb) (n=3)	Passive (including: adjectival & stative passive (n=6)	Causative & Causative-passive (n=2)
Break	Freeze	Please	Kill (be killed)	Wash (make X
Wash	Fall	Confuse	Break (be broken)	wash Y)
Kill	Fall from	Shock	Close (be closed)	Work (be
Close			Confuse (be confused)	made to work)
Stop			Interest (be interested)	
			Surprise (be surprised)	

Verbs were mostly selected from the first (most frequently used) vocabulary band⁶ i.e. 1 to 1,000 for English while a couple of them, *shock* and *confuse*, are in the second band. In some cases, the informant’s ability to use the same verb in canonical and non-canonical constructions was tested; for example, the verb *kill* was included both in an active context and a passive one.

6. English frequency list based on Vp-BNC list (http://www.lex Tutor.ca/freq/lists_download/1000_families.txt)

7. Results

7.1 Vocabulary size test

Figure 3 lists the distribution of the vocabulary size for 22 informants. Minimum and maximum sizes are 3,000 and 12,700 word families respectively (mean 7,141; $SD=2,466$). Nation & Beglar note that “undergraduate non-native speakers successfully coping with study at an English speaking university have a vocabulary around 5,000–6,000 word families. Non-native speaking Ph.D. students have around a 9,000 word vocabulary” (2007, p. 9). Out of the 22 informants in the present study, ranging from well below undergraduate university level to beyond Ph.D. level, three informants were selected for a focused investigation of their syntactic development in ESL from each lexical size group: Top, Middle and Low. Their vocabulary size and other relevant information are summarized in Table 5. These nine informants were asked to proceed to perform the profiling and translation tasks. The last column in the table lists the total number of turns each informant produced through the profiling task.

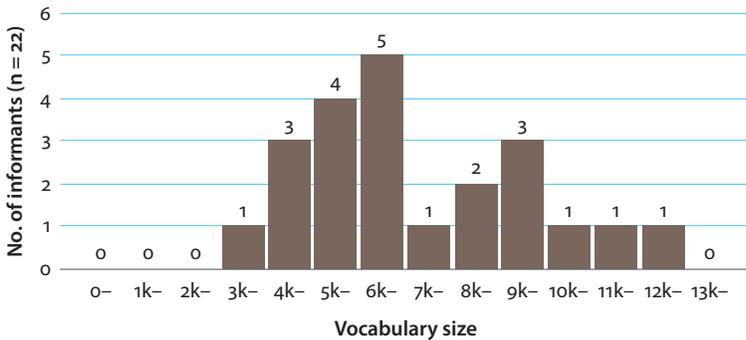


Figure 3. 22 informants’ lexical size

7.2 Profiling task: Question sentence constructions

This section presents the analysis of the informant’s production of question sentences produced through the profiling task. Table 6 summarises frequency count of their question sentences production according to the question type: (1) yes/no questions and (2) wh-questions while Table 7 shows the breakdown of the question sentence production against PT stages. A particular syntactic stage is considered to be acquired in PT when an informant produces any construction belonging to that stage more than once with lexical variation (this excludes formulaic or echoic production). Applying this acquisition criterion, the nine informants’ PT stages in question sentences are identified. In Table 7, the highest stage acquired by each

Table 5. Lexical size and background information of High, Mid and Low vocabulary size learners

Group	Code Name (Male or Female)	Vocabulary Size	Age	Length of stay in Australia (current occupation)	Total no. of turns produced via Profiling task
High	JA 03 (F)	12,700	43	8 years (Translator)	120 turns
	JA 13 (F)	11,200	29	2yrs & 9 months (MA student)	96
	JA 02 (M)	10,100	27	2yrs & 6 months (Ph.D. student)	150
Mid	JA 06 (F)	6,900	32	4 months (Wife of an engineer sent to Australia for business)	104
	JA21 (F)	6,800	32	9 months (Employee at a Japanese agency)	97
	JA 08 (F)	6,800	21	6 months (Working holiday participant)	119
Low	JA 19 (F)	4,600	24	8 weeks (Student of an English school)	146
	JA 20 (F)	4,100	34	4 weeks (Student of a short vocational course)	182
	JA 11 (F)	3,000	40	6 months (House wife)	130

informant is shaded: J11 and J19 are at Stage 1; J20 is at Stage 3; all the others J08, J21, J06, J02, J13 and J03 are at Stage 4. Note that “-” (minus) next to the number as in “-1” indicates negative evidence for acquisition. Notice that numbers are listed in the cell only if the informant produced a particular structure.

Table 6. Summary on the frequency of question sentence constructions by the nine informants

Group	Low			Mid			High		
	J11 (3.0)	J20 (4.1)	J19 (4.6)	J08 (6.8)	J21 (6.8)	J06 (6.9)	J02 (10.1)	J13 (11.2)	J03 (12.7)
Informants (lexical size: x 1,000)									
Y/N questions	6	7	2	16	9	7	7	5	6
Wh-questions	1	5	6	6	3	10	5	7	7
Total	7	12	8	22	12	17	12	12	13

Table 7. Breakdown of the question sentences against PT stages

Stage	Structure	Low			Mid			High			
		J11 3.0	J19 4.6	J20 4.1	J08 6.8	J21 6.8	J06 6.9	J02 10.1	J13 11.2	J03 12.7	
4	marked alignment	Y/N questions	1/-1		1	7	2	2	1	1	
		Wh-questions	0	1/-1	-1	2	2	8	1	5	4
3	QUE particle + unmarked alignment	Y/N questions	1	1	5/-1	3		4		4	1
		Wh-questions			2						
2	unmarked alignment	Y/N questions	1			2	1		1		
		Wh-questions									
1	lemma access	Y/N questions	2	1		11	1	1	4		4
		Wh-questions	1	4	2	4	1	2	4	2	3

7.2.1 Low vocabulary size informants

Two of three Low vocabulary size informants (J11 and J19) are still at the lowest Stage 1 (i.e., Lemma access) because they used only single words or formulaic questions as in (7) and (8). One Low vocabulary size informant, J20, showed substantial evidence for Stage 3 QUE + Unmarked alignment as in (9).

(7) J11 Black cats?

(8) J19 when?

(9) J20 does girl have ball ball?

Also, participants from the Low vocabulary size group were often unable to complete the question sentences; they started with a particular question construction but changed the question sentence structure from Wh to Y/N or the other way around. In the example (10 a), J20 started a “do-question” (Stage 3, QUE [Unmarked alignment]) but was unable to complete it. J20, instead, changed the pattern and attempted “wh-question” requiring Stage 4 operation (Marked alignment) but was unsuccessful. They also have problems selecting the appropriate auxiliary verb to form a question as in (10 b).

(10) a. J 20 do you have (X) do you.? hum hummm how many oh no no no
no uhm how many people. there is. there the. bench in ? (laugh)

b. J 19 how long how long uhm di do do you uh how long are you there?

Participants from the Low vocabulary size group also attempted a higher stage structure than had been reached and ended up ungrammatical, i.e., negative

evidence of acquisition. The sentence (11) exemplifies an unsuccessful attempt of Stage 4 (Marked Alignment), which lacks the subject.

(11) J11 **can see cats?*

Further, these two learners showed some problems with Wh-questions involving SUBJ as in (12) and (13). These examples exhibit incorrect functional assignment of the WH-pronoun by providing an extra SUBJ in the sentence. Functional assignment requires the procedural skill placed at Stage 4 according to PT but these informants have not attained that stage.

(12) J19 **how many birds are you here?*

(13) J20 **how many people di did did did you ride your (X) ride a bicycle?*

7.2.2 *Mid vocabulary size informants*

All of the mid vocabulary size informants have attained Stage 4 (Marked alignment). (14) and (15) are examples of Stage 4 WH and (16) Y/N question which require marked alignment.

(14) J08 *ok. ah. how long have you been?*

(15) J06 *umm ok. which language do you usually use?*

(16) J21 *can you see one spider in the middle of this?*

Although these informants are at Stage 4, they also produced the lower, Stage 3, unmarked alignment Y/N question simply using raising intonation as in (17). J21 and J08 produced such question sentences once and twice respectively.

(17) J08 *brother is ah special school. school?*

Another observation among the Mid vocabulary size informants is that J08 (lexical size 6.8) lacks production of Stage 4 Y/N questions.

7.2.3 *High vocabulary size informants*

All three high vocabulary size informants produced a variety of question sentences belonging to different stages. All of them are at Stage 4 and produced both Y/N and WH-questions at this stage. There was no ungrammatical production among these informants. Unlike Low vocabulary size informants, J02 was able to produce WH-question asking SUBJ information correctly as follows in (18). Superficially, this SUBJ in-situ question sentence follows unmarked alignment. But without acquiring functional assignment of the event participants, production of this sentence pattern is not possible. In conclusion, two Low vocabulary size informants (size 3.0k and 4.6k) were at Question Stage 1 and one (size 4.1k) at Stage 3. All Mid and High vocabulary size informants are at Stage 4.

(18) J02 *who is trying to feed duck?*

7.3 Translation task: Argument-Grammatical function mapping

This section presents the results of the translation task which investigates the informant's ability to map thematic roles (e.g., Agent, Patient, Location) on to grammatical functions (e.g., SUBJ, OBJ). Table 8.a shows acquisition of lexically non-default mapping while Table 8.b presents structurally non-default mapping.

Table 8.a Default versus lexically non-default mapping

Vocab size Group	Informant	Vocab. size (x1,000)	Default	Non-default	
				Unaccusative	Psych Verb
Low	JA19	4.6	3/5 (.6)	1/3 (.33)	0/3
	JA11	3.0	4/5 (.8)	1/3 (.33)	0/3
	JA20	4.1	5/5 (1.0)	3/3 (1.0)	0/3 (0)
Mid	JA08	6.8	4/5 (.8)	2/3 (.67)	0/3 (0)
	JA06	6.9	5/5 (1.0)	3/3 (1.0)	0/3 (0)
	JA21	6.8	5/5 (1.0)	2/3 (.67)	2/3 (.67)
High	JA02	10.1	5/5 (1.0)	3/3 (1.0)	3/3 (1.0)
	JA13	11.2	5/5 (1.0)	3/3 (1.0)	3/3 (1.0)
	JA03	12.7	5/5 (1.0)	3/3 (1.0)	3/3 (1.0)
	NS control	11.3	5/5 (1.0)	3/3 (1.0)	3/3 (1.0)

Table 8.b. Default versus structurally non-default mapping

Vocab size Group	Informant	Vocab size (x1,000)	Default	Non-default	
				Passive	Causative
Low	JA19	4.6	3/5 (.6)	1/6 (.17)	0/2 (0)
	JA11	3.0	4/5 (.8)	1/6 (.17)	0/2 (0)
	JA20	4.1	5/5 (1.0)	2/6 (.33)	0/2 (0)
Mid	JA08	6.8	4/5 (.8)	2/6 (.33)*	1/2 (.5)
	JA06	6.9	5/5 (1.0)	4/6 (.67)	1/2 (.5)
	JA21	6.8	5/5 (1.0)	5/6 (.83)	2/2 (1.0)
High	JA02	10.1	5/5 (1.0)	6/6 (1.0)	2/2 (1.0)
	JA13	11.2	5/5 (1.0)	5/6 (.83)	2/2 (1.0)
	JA03	12.7	5/5 (1.0)	6/6 (1.0)	2/2 (1.0)
	NS control	11.3	5/5 (1.0)	5/5 (1.0)	2/2 (1.0)

(The numbers before and after the slash indicate the number of correct instances and the number of contexts respectively. The numbers in brackets represent accuracy rates.)

In these tables, the informants are organised vertically against acquisition of default and non-default mapping. The cell is shaded when the informant produced the structure more than once with some lexical/structural variety. Note the PT stages of “lemma access” and “Default Mapping plus extra argument” are not listed in these tables. Also, acquisition of default mapping appears in both tables for comparison with each type of non-default mapping.

7.3.1 *Default mapping*

All nine informants show enough evidence of acquisition of default mapping in the translation task. However, two Low vocabulary size informants, JA 19 and JA11, and one Mid vocabulary size informant, JA 08, made some errors in the construction. Interestingly these errors involve the use of “be+past participial” as in (19) and (20) instead of canonical transitive/intransitive. All other informants successfully produced default mapping.

(19) J19 *Porice was stoped that car.⁷
(source: 警察は、その車を止めた。 “The police stopped that car”)

(20) J08 *My dog was broken the daughter’s doll
(source: 私の犬が娘の人形をこわした “My dog broke my daughter’s doll”)

7.3.2 *Lexically non-default mapping*

Default and non-default mapping with unaccusative and psych verbs showed interesting implicational patterns. Two Low vocabulary size learners, JA 19 and JA11 had problems in mapping argument roles correctly with unaccusative verbs. Errors in mappings are exemplified in (21a–b). It is interesting to see in (21b) JA11 attempted canonical mapping SVO but this unaccusative verb, *fall*, does not take an Agent: the subject position (i.e., preverbal position) is left empty and the Theme *tree* is placed in the post verbal position. On the other hand, all three High informants were able to produce sentences with unaccusative verbs with target-like mapping as in (21c).

- (21) Translation of 庭の木が倒れた。 into English
- JA 19 *The tree was falled in my girden.
 - JA 11 *fall in down gerden tree.
 - JA03 A tree in our yard fell.

The structures involving psych verbs are acquired later than unaccusative verbs that require lexically non-default mapping. All Low vocabulary size informants

7. All translations by the informants are presented without correcting spelling errors.

and two Mid informants (i.e., JA08 and JA06) were unable to construct non-default mapping with psych verbs at all. See examples (22) and (23) below.

- (22) J19 The plain accident was shocked to around the world.
 (source: その飛行機事故は、世界中の人にショックをあたえた。
 ‘The airplane accident shocked people all over the world.)
- (23) J08 Yamada teacher is confused for students because for his explanation.
 (source: 山田先生の説明は、いつも学生を混乱させる。
 ‘Professor Yamada’s explanation always confuses his students.’)

In these examples, the informants encoded the event participants in their translation into English sentences in the same order as the source sentences in Japanese: *the plane accident* > (*people*) *around the world* in (22); *Yamada* > *the students* in (23) leading to false mapping between argument roles and grammatical functions. Only the High vocabulary size group informants were able to perform target-like mapping operations with the sentences involving both unaccusative and psych verbs. This finding is consistent with White, et. al. (1998) which report that L2 learners have problems with Experiencer-Object psych verbs and that Lower intermediate learners of English made more errors than participants in the high Intermediate category.

7.3.3 Structurally non-default mapping

Similar to lexically non-default mapping, the informants acquired structurally non-default mapping later than default mapping. As seen in Table 8.b., (on production of default mapping, passive and causative forms implicational pattern) those informants who are able to produce target-like mapping with passive are also able to do it with default mapping; those informant who are able to produce correct mapping with Causative are also able to do so with passive but not the other way around. The following examples, (24a) passive and (25a) causative, show that Lower vocabulary size informants seem to have problems creating structural frames encoding non-default argument mapping. On the other hand, Higher vocabulary size learners had no problem with passives and causatives as in (24b) and (25b).

- (24) Translation of トムは、メアリーに殺された。‘Tom was killed by Mary’ into English
 a. J11 *Tom killed by Mary
 b. J21 Tom was killed by Mary.
- (25) Translation of 母は、毎日私にお皿を洗わせる。‘My mother made me wash the dishes every day’
 a. J20 *I am washed dishes by mother everyday
 b. J03 My mother makes me wash the dishes every day.

8. Discussion

Now we turn to discussing the relationship between vocabulary size and the acquisition of question construction and argument mapping based on the results gained through this study. This section aims to answer to the two research questions proposed in the introductory section above.

8.1 Relationship between vocabulary size and acquisition of question construction

Figure 4 shows that PT stages in question constructions, as achieved by the nine informants, are related to their vocabulary sizes. The two Low vocabulary size learners J11 and J19 are at the lowest PT stage and one Low, J20 at Stage 3. All Mid and High vocabulary size learners are at the highest PT stage. According to Nation and Beglar (2007), the overseas students who are able to cope with undergraduate study successfully at English medium universities possess around 5,000–6,000 vocabulary size. Mid and High vocabulary size learners in the current study are beyond this level as all of them achieved the highest PT stage. This indicates that ESL learners of Mid and High vocabulary size are able to produce both Y/N and Wh-questions including marked alignment without any problem. This is not true for Low vocabulary learners.

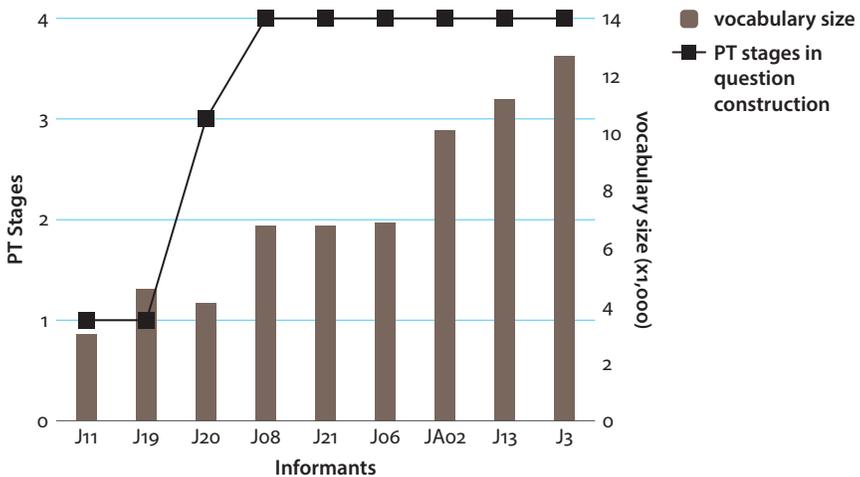


Figure 4. Informant's vocabulary size and their PT stages in question constructions

8.2 Relationship between vocabulary size and acquisition of argument mapping

Since Fig 4 clearly shows the relationship between vocabulary size and the acquisition of questions and given that mapping is currently divided into two broad

stages (default and non-default), I will show the level of accuracy in mapping, rather than look only at stages. This will allow for a more fine-grained relationship between vocabulary size and mapping ability of the more advanced learners. Figure 5 shows relationship between vocabulary size and acquisition of default mapping while Figure 6 presents the relationship between vocabulary size and acquisition of non-default mapping. Figure 5 shows a predictably high level of accuracy in default mapping. Thus, six out of nine informants achieved the maximum accuracy rate 1.0. But also the three informants, two from Low and one from Mid vocabulary size groups, who made some errors, nevertheless have high accuracy rates with default mapping (0.6 or above). Therefore, it can be said that default mapping is acquired early.

In Figure 6 we can see that measuring accuracy rates of various non-default mappings will differentiate more accurately between intermediate and more advanced learners because different vocabulary size groups (Low, Mid and High) showed distinctively different performance patterns. Only the High vocabulary size group was able to attain about 100% accuracy with all types of non-default mapping. In contrast, participants from the Low group had great difficulties with non-default mapping: especially causative and psych verb construction were not achieved at all while passive and unaccusative constructions showed low accuracy rates. As for the Mid group, although their three vocabulary sizes are very close to each other (i.e., 6,800, 6,900 and 6,900), their performance patterns are not uniform. JA08 is close to the performance of the Low group while JA 21 is close to High group. Notice that two Mid informants, JA08 and JA06 were unable to create causative constructions. Thus, acquisition of non-default mapping is a key indicator even for learners with vocabulary size 6,000 or above.

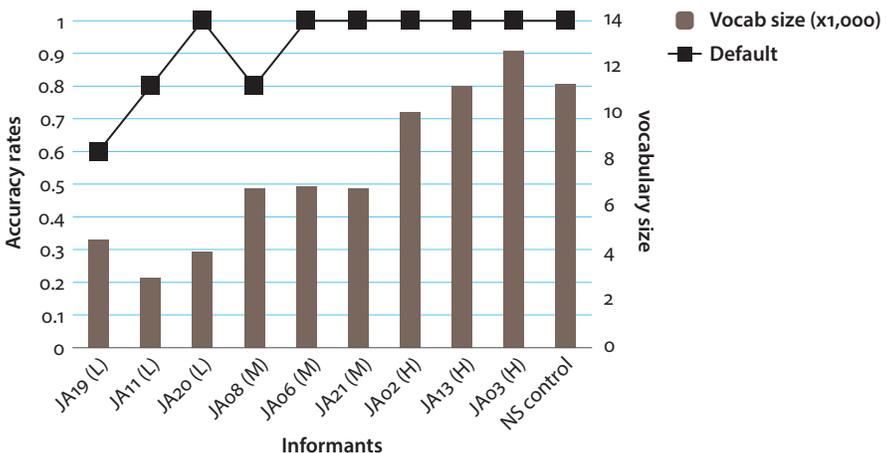


Figure 5. Vocabulary size and default mapping

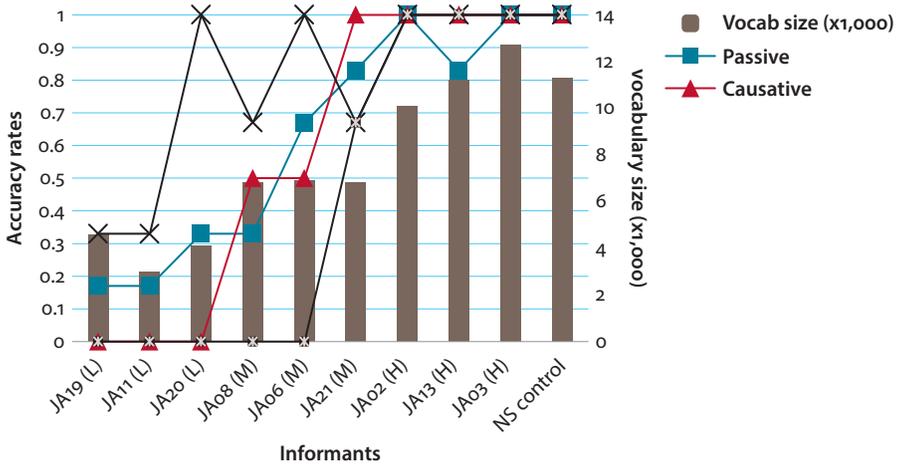


Figure 6. Vocabulary size and various non-default mapping

9. Conclusion

This study within the framework of Processability Theory investigated the relationship between vocabulary learning and syntactic development in English as a second language. Question sentences and default/non-default mapping were the focus of our syntactic analysis. The data analysis on question sentences was conducted based on Bettoni & Di Biase (2015) which incorporated TOP and FOC following Darlymple's (2001) LFG interpretation of question sentences. Broadly speaking (at least for TOP) this was part of the 2005 PT extension.

The findings can be summarized as follows. Regarding question sentence constructions (Research Question 1), Mid and High vocabulary sizes (6,000 or over) can predict highest L2 developmental stage as defined by PT (i.e., marked alignment). All Low group learners had difficulties in constructing various English question sentences. The problems include marked alignment, selection of auxiliary verbs, and constructions of SUBJ Wh-questions. As for argument mapping (Research Question 2), the acquisition of default and non-default mapping showed implicational relations: all types of non-default mappings are acquired only after default mapping is in place. This relation was observed with both lexically and structurally non-default mappings. Within lexically non-default mapping, both Low and Mid groups were unable to construct correct mapping with psych verbs. Within structurally non-default mapping, the Low group showed problems with both passive and causative forms. The Mid group also showed problems with non-default mapping but their performances are distinctively better than the Low group.

The High group learners were able to cope with all types of non-default mapping. Productive ability with non-default mapping then seems to be an accurate indicator of syntactic ability but it also presupposes a high vocabulary size. From the learning and teaching point of view a clear awareness of the importance of vocabulary size and non-default mapping helps plan more focused interventions to promote further language development. In this study, I attempted to use a translation task to elicit sentences involving transitive/intransitive contrast. The use of such task may involve some problem in PT studies if they were done exclusively with translation tasks. However, production elicited via more 'traditional' spot-the-difference tasks and free conversation ensure normal profiling. If anything, the use of translation tasks augments the range of methodologies available. Although appropriateness of the use of translation task in PT may not be conclusive from these studies, the results seem to indicate that less formal writing (without editing) follows the PT schedule. The translation task was conducted with pen and paper without eraser and may be considered close to online language production (in any case the L2 informant's writing including their editorial changes on the translation can be traced). I believe that it is worthwhile exploring the possibility of using translation task in PT studies especially because it opens up a different modality and it can easily be controlled by parallel oral production tasks, as I do in this study.

Acknowledgment

This study was funded by the University of Western Sydney (UWS) School Research Funding 2010–2011. An earlier version of this paper was presented at the Symposium on the Bilingualism and Intercultural Communication, August, 2013, School of Humanities & Communication Arts, UWS. I thank the audience at the Symposium for their invaluable comments and advice to improve this paper. I am also thankful of Bruno Di Biase and two anonymous reviewers for their helpful comments.

References

- Bettoni, C., & Di Biase, B. (Eds.). (2015). *Processability Theory: Current issues in theory and application*. European Second Language Association.
- Birrel, B. (2006). Implications of English standards among overseas students at Australian universities. *People and Place*, 14(4), 53–64.
- Bock, K., & Levelt, W. J. M. (1994). Language production: Grammatical encoding. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics*. San Diego, CA: Academic Press.
- Bresnan, J. (2001). *Lexical-functional syntax*. Malden, MA: Blackwell.
- Burzio, L. (1986). *Italian Syntax: A Government-Binding Approach*. Dordrecht: Reidel.
- doi:10.1007/978-94-009-4522-7

- Choi, H.-W. (2001). Phrase structure, information structure, and resolution of mismatch. In P. Sells (Ed.), *Formal and empirical issues in optimality theoretic syntax* (pp. 17–62). Stanford, CA: CSLI.
- Comrie, B., Matthews, S., & Polinsky, M. (2003). *The atlas of languages* (rev. ed.). London: Euro-span Group.
- Culicover, P., & Jackendoff, R. (2005). *Simpler Syntax*. Oxford: Oxford University Press.
doi:10.1093/acprof:oso/9780199271092.001.0001
- Dalrymple, M. (2001). *Syntax and semantics. Lexical functional grammar. Vol. 34*. San Diego, SA: Academic Press. doi:10.1163/9781849500104
- DeKeyser, R. (2007). Skill acquisition theory. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition* (pp. 97–135). Mahwah, NJ: Laurence Erlbaum Associates.
- Di Biase, B., Kawaguchi, S., & Yamaguchi, Y. (2015). English. In Bettoni & Di Biase (Eds.) (pp. 85–115).
- Foley, W. A., & Van Valin, R. D. (1984). *Functional syntax and universal grammar*. Cambridge: Cambridge University Press.
- Givón, T. (1984). *Syntax: A functional-typological introduction*. Amsterdam: John Benjamins.
doi:10.1075/z.17
- Greenberg, J. H. (1966). *Universals of language*. Cambridge MA: The MIT Press.
- Håkansson, G., & Norrby, C. (2007). Processability Theory applied to written and oral Swedish. In F. Mansouri (Ed.), *Second language acquisition research: Theory-construction and testing* (pp. 81–94). Newcastle upon Tyne: Cambridge Scholars.
- Harmer, J. (2007). *The practice of English Language Teaching*. Harlow: Pearson.
- Harrington, M., & Carey, M. (2009). The on-line Yes/No test as a placement tool. *System*, 37(4), 614–626. doi:10.1016/j.system.2009.09.006
- Hirakawa, M. (2003). *Unaccusativity in second language Japanese and English*. Tokyo: Hituzi Shobo.
- Hopp, H. (2007). *Cross-linguistic differences at the syntax-discourse interface in off- and on-line L2 performance*. Paper presented at the Proceedings of the 2nd Conference on Generative Approach to Language Acquisition North America (GALANA), Somerville, MA.
- Jackendoff, R. S. (1972). *Semantic Interpretation in Generative Grammar*. Cambridge, MA: The MIT Press.
- Kaplan, R. M., & Bresnan, J. (1982). Lexical-Functional Grammar: A formal system for grammatical representation. In J. Bresnan (Ed.), *The mental representation of grammatical relations* (pp. 173–281). Cambridge, MA: The MIT Press.
- Kawaguchi, S. (2009). Acquiring causative constructions in Japanese as a second language. *The Journal of Japanese Studies*, 29(2), 273–291.
doi:10.1080/10371390903066657
- Kawaguchi, S. (2013). The relationship between lexical and syntactic development in English as a second language. In A. F. Mattsson & C. Norrby (Eds.), *Language Acquisition and Use in Multilingual Contexts: Theory and Practice* (Vol. 52, pp. 92–106). Lund: Travaux de L'institut de Linguistique de Lund.
- Kawaguchi, S. (2015). Connecting CALL and second language development: e-tandem learning of Japanese. In Bettoni & Di Biase (Eds.) (pp. 291–305).
- Keenan, E., & Comrie, B. (1977). Noun phrase accessibility and universal grammar. *Linguistic Inquiry* 8(1), 63–99.
- Kempen, G., & Hoenkamp, E. (1987). An incremental procedural grammar for sentence formulation. *Cognitive Science*, 11(2), 201–258. doi:10.1207/s15516709cog1102_5

- Kroll, J. F., & Tokowicz, N. (2001). The development of conceptual representation for words in a second language. In J. L. Nicol (Ed.), *One mind, two languages: Bilingual language processing. Explaining linguistics* (pp. 49–71). Malden, MA: Blackwell.
- Laufer, B., & Hulstijn, J. (2001). Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied Linguistics*, 22(1), 1–26.
doi:10.1093/applin/22.1.1
- Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: The MIT Press.
- Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22(1), 1–38.
- Levin, B., & Rappaport Hovav, M. (2005). *Argument realization*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511610479
- Mochida, A., & Harrington, M. (2006). The Yes/No test as a measure of receptive vocabulary knowledge. *Language Testing*, 23(1), 73–98. doi:10.1191/0265532206lt3210a
- Nation, I. S. (2006). How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review/La Revue Canadienne des Langues Vivantes*, 63(1), 59–82.
doi:10.3138/cmlr.63.1.59
- Nation, I. S. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press. doi:10.1017/CBO9781139524759
- Nation, P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher*, 31(7), 9–13.
- Payne, T. E. (2011). *Understanding English Grammar*. Cambridge: Cambridge University Press.
- Pienemann, M. (1998). *Language processing and second language development: Processability Theory*. Amsterdam: John Benjamins. doi:10.1075/sibil.15
- Pienemann, M., & Kessler, J.-U. (Eds.). (2011). *Studying Processability Theory: An introductory textbook*. Amsterdam: John Benjamins. doi:10.1075/palart.1
- Pienemann, M., Di Biase, B., & Kawaguchi, S. (2005). Extending Processability theory. In M. Pienemann (Ed.), *Cross-linguistic aspects of Processability Theory* (pp. 199–252). Amsterdam: John Benjamins. doi:10.1075/sibil.30.09pie
- Pinker, S. (1984). *Language learnability and language development*. Cambridge, MA: Harvard University Press.
- Rahkonen, M., & Håkansson, G. (2008). Production of written L2-Swedish-processability or input frequencies. In J.-U. Kessler (Ed.), *Processability approaches to second language development and second language learning* (pp. 135–161). Newcastle upon Tyne: Cambridge Scholars.
- Slobin, D. I. (1982). Universal and particular in the acquisition of language. In E. Wanner & L. R. Gleitman (Eds.), *Language acquisition: State of the art* (pp. 128–170). Cambridge: Cambridge University Press.
- Sorace, A. (2003). Near-nativeness. In C. Daughy & M. Long (Eds.), *Handbook of second language acquisition theory and research* (pp. 130–151). Oxford: Blackwell.
doi:10.1002/9780470756492.ch6
- VanPatten, B., & Houston, T. (1998). Contextual effects in processing L2 input sentences. *Spanish Applied Linguistics*, 2(1), 53–70.
- Van Valin, R. D. (2005). *Exploring the syntax-semantics interface*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511610578
- Van Valin, R. D., & LaPolla, R. J. (1997). *Syntax: Structure, meaning, and function*. Cambridge: Cambridge University Press. doi:10.1017/CBO9781139166799
- White, L., Brown, C., Bruhn-Garavito, J., Chen, D., Hirakawa, M., & Monrul, S. (1998). Psych verbs in second language acquisition. In E. C. Klein & G. Martohardjono (Eds.), *The development of second language grammars: A generative approach* (pp. 171–196). Amsterdam: John Benjamins. doi:10.1075/lald.18.1owhi

- Yamaguchi, Y. (2008). The early development in child L2 acquisition: What happens after “canonical order”? In J.-U. Keßler (Ed.), *Processability approaches to second language development and second language learning* (pp. 245–266). Newcastle upon Tyne: Cambridge Scholars.
- Yamaguchi, Y. (2009). The development of plural marking and plural agreement in child English L2 acquisition. In J.-U. Keßler & D. Keatinge (Eds.), *Research in second language acquisition: Empirical evidence across languages* (pp. 9–39). Newcastle upon Tyne: Cambridge Scholars.
- Yamaguchi, Y. (2010). *The acquisition of English as a second language by a Japanese primary school child: A Longitudinal study from a Processability Viewpoint*. Unpublished Ph.D. dissertation, University of Western Sydney.
- Zhang, Y., & Mi, Y. (2010). Another look at language difficulties of international students. *Journal of Studies in International Education* 14, 371–388. doi:10.1177/1028315309336031

Appendix

Translation task used in the study with example answers by JA03 (vocabulary size 12,700) and J11 (vocabulary size 3,000)

	日本語の文	動詞	英語の文
1	わたしは、オーストラリアの映画に興味がある。	interest	JA03: I am interested in Australian movies. JA11: I am interesting for Australian movie.
2	私の犬が娘の人形をこわした。	break	JA03: My dog broke my doll. JA11: My dog broke my dou-tear's doll.
3	私が料理をつくる。	cook	JA03: I cook. JA11: I cook.
4	主人がお皿を洗う。	wash	JA03: My husband washes dishes. JA11: My husband washed dishes.
5	テストの結果を見て、私はびっくりした。	surprise	JA03: I was surprised to see the exam results. JA11: I surprised about my test.
6	山本さんの猫が、私の鳥を殺した。	kill	JA03: Yamamoto's cat killed my bird. JA11: My bird killed by Yamamoto's cat.
7	けい子は、博史にプレゼントをもらった。	receive	JA03: Keiko received a gift from Hiroshi. JA11: Keiko received for Hiroshi's presents.
8	私たちは、この事故を警察に届けなければならない。	report	JA 03: We must report this accident to the police. JA11: We have to take a report to porice satesion.
9	私は、いつも店のドアを7時にしめる。	close	JA03: I always close the door of my shop at 7. JA11: I close the door at 7.
10	猫が木から落ちた。	fall	JA03: A cat fell off the tree. JA11: Cat's fall down by tree.

(Continued)

	日本語の文	動詞	英語の文
11	山田さんは、みんなに旅行の写真を見せた。	show	JA03: Yamada-san showed everyone his photos from his travelling. JA11: Mr Yamada show us his trip picture.
12	この時計は、高そうだ。	seem	JA03: This clock seems expensive. JA11: This watch seems like expensive.
13	トムのプレゼントは、たいへん私を喜ばせた。	please	JA03: I was very pleased with Tom's gift. JA11: I pleased by Tom's present.
14	この店のドアは、いつも閉まっている。	close	JA03: The door to the shop is always closed. JA11: door's closed that store.
15	そのニュースを聞いて、私は、とても混乱した。	confuse	JA03: I was very confused after hearing the news. JA11: I confused about that news.
16	この時計は、壊れている。	break	JA03: This watch is broken. JA11: This watch broke already.
17	庭の木が倒れた。	fall	JA03: A tree in our yard fell. JA11: fall in down garden tree.
18	母は、毎日私にお皿を洗わせる。	wash	JA03: My mother makes me wash the dishes every day. JA11: my mother
19	山田先生の説明は、いつも学生を混乱させる。	confuse	JA03: Professor Yamada's explanation always confuses his students. JA11: Mr Yamada make confused his students.
20	トムは、メアリーに殺された。	kill	JA03: Tom was killed by Mary. JA11: Tom killed by Mary.
21	わたしは、息子が大学に合格すると信じている。	believe pass	JA03: I believe that my son will pass the university entrance exam. JA11: I believe my son passed by University.
22	水は、0度で凍る。	freeze	JA03: Water freezes at 0 degree. JA11: Water freeze 0°.
23	警察は、その車を止めた。	stop	JA03: The police stopped the car. JA11: Police stopped the car.
24	その飛行機事故は、世界中の人にショックをあたえた。	shock	JA03: The airplane accident shocked people all over the world. JA11: That airplane accident make would people shocked.
25	わたしは、ボスに毎日8時まで仕事をさせられる。	work	JA03: I am made to work until 8 by my boss every day. JA11: I had work at 8 every day by my boss.

Processability Theory and language development in children with Specific Language Impairment

Gisela Håkansson
Lund University

Children with Specific Language Impairment (SLI) represent a special group among young monolingual children, since they have problems acquiring their first language. Most research deals with English-speaking children, and points to bound morphology as the problematic area. However, cross-linguistic studies show that SLI characteristics differ between languages, and that it is not always bound morphology that is affected but sometimes other phenomena, for example syntax or function words. The seemingly contradictory findings can be accommodated within Processability Theory (PT) and from the point of view of feature unification at different levels of processability. Focussing on individual performances instead of group means changes the perspective and makes it possible to analyze children with SLI as learners along a developmental continuum.

1. Introduction

Not all children develop their first language in an unproblematic way, but some exhibit considerable difficulties in phonological, grammatical, lexical and/or pragmatic aspects of language. Approximately 5–7 % of all children are diagnosed with Specific Language Impairment (SLI), most often with some grammatical problems. The two main questions within research on SLI are:

1. What is the problem; is it a representational deficit or an auditory processing deficit?
2. Are there specific structures that are “vulnerable”, i.e. likely to be affected?

The traditional method in SLI research is to analyse production or comprehension data from children diagnosed with SLI and compare it to two control groups. One group consists of peers of the same age (age-matched) and the other of children who are younger but have the same utterance length measured in Mean Length of Utterance (MLU match). The most interesting comparison is that between younger

children with the same utterance length as the children with SLI. If they have the same MLU, their grammar is expected to be similar, and the difference between children with SLI and children with TD can be described as a case of language delay. If the children with SLI have the same length of utterance but different features of grammar, the difference cannot be only a matter of delay, but of deviance. Those structures that are not similar are assumed to be clinical markers of SLI.

Bound morphology, particularly on verbs, has long been assumed to be the critical feature (for example: “There seems to be a consensus that SLI children have problems in the area of grammatical morphology”. Clahsen 1992:3). However, empirical studies from different languages have shown that there are “striking cross-linguistic differences among children with SLI” (Leonard 2009:169), and what is typical for one language is not necessarily typical for SLI in another language. For English-speaking children, the most common problem is verb morphology (Rice & Wexler 1996). For Swedish children however, the problem is not morphology but word order, that is, subject-verb inversion (Håkansson 1997, 2001; Hansson, Nettelblatt & Leonard 2000), whereas the typical feature in Italian and Spanish SLI is neither verb morphology nor word order, but function words such as articles and object clitics (Bedore & Leonard 2005).

In this paper I will argue that the problems for children with SLI rest neither on the morphological, syntactic nor lexical surface structures *per se*, but on the requirements of grammatical processability that underly them. The idea is that children with SLI are language learners and therefore can be described within a framework of language development. The non language-specific nature of Processability Theory (Pienemann 1998, 2005) makes it highly suitable to deal with language impairment in different languages. Due to typological differences between languages, language impairment may surface in what is traditionally seen as different linguistic domains. This is the reason why in some languages morphology seems to be impaired, but in other languages syntax or lexical items are affected.

The paper is organized in the following way. First, a short overview of earlier research on SLI will be given. Then I will present a reanalysis of data from the project “Grammatical processes in language acquisition” (Håkansson 1997, 2001; Håkansson & Hansson 2000). Finally, the potential of PT as a model to explain cross-linguistic differences in SLI will be discussed.

2. Earlier research

2.1 What is the problem – representation or processing?

Much research on SLI has been devoted to finding causal factors. The search has been for either linguistic or perceptual factors to explain the problem – the

underlying assumption being that SLI is a syndrome with one single cause. Thus, there are two main perspectives: one characterizes SLI as a representational problem, the other as an auditory processing problem.

Within the first perspective, one early proposal of a domain-specific deficit in the linguistic representation is the *feature blindness hypothesis* (Gopnik 1994), which assumes that grammatical features such as person and tense are missing from the underlying grammars. Another example is the *missing agreement account*, according to which children with SLI have problems in establishing agreement relationships, such as using subject-verb agreement markings (Clahsen 1992). The *extended optional infinitive hypothesis* (EOI), which is formulated by Rice and Wexler (1996), claims that children with SLI have an extended period where finiteness is only optionally marked. Typically developing children also have a period of optional infinitives, but it is short-lived and soon disappears. The suggestion is that there is a biologically determined program for optional infinitives and children with SLI have a deficit in this biological program. Each of these three hypotheses refers to problems encountered with the production of verb morphology in English-speaking children. A fourth proposal for a domain-specific gap in the SLI grammar is that there are problems in establishing hierarchical relationships between linguistic structures. Here, the data come from comprehension experiments that are assumed to reflect internal grammatical representations. English-speaking children with SLI show difficulties in comprehending structures with dependent relationships (e.g. reflexives and passives). Van der Lely (1998) claims that, at least for a subtype of SLI, the problem is a *Representational Deficit for Dependent Relationships* (RDDR), which means that the children are unable to link grammatical features.

The second main perspective focusses on auditory processing and suggests that the children's problems are due to difficulties with the processing of linguistic input. The *surface hypothesis* (Leonard 1989, 1998) claims that children with SLI are limited in their auditory processing capacity. They have, for example, difficulties in processing and producing unstressed syllables and morphemes of short duration. From this perspective, the well-known problem for English-speaking children with SLI in producing 3rd person singular -s is interpreted as a difficulty with the processing of morphemes of short duration (Leonard 1998). The finding that Italian-speaking children do not exhibit problems with verb inflections seems to be in accordance with this hypothesis, since Italian verb inflections are syllabic and of longer duration than their English counter-parts. The fact that the same English morpheme, the suffix -s, is used by children with SLI as plural marker on nouns but not as person marker on verbs, has been explained by reference to duration and frequency in the input (Hsieh, Leonard, & Swanson 1999). Plural nouns appear in sentence-final position more often than third person singular verb forms, and are

therefore more likely to be of longer duration. Furthermore, frequency in the input may also be beneficial for children with poor auditory processing skills, as plural -s is much more frequent than third singular verb -s in the input to the children, both in caretaker's speech and in story books (Hsieh et al. 1999).

2.2 Are there specific structures that are likely to be affected?

There are two main reasons behind the intense search for vulnerable structures. One is the practical goal to find suitable tools for language assessment in different languages. The other is the existence or non-existence of vulnerable structures, which would have strong implications for the definition of SLI. In Leonard (1998:66) it is stated that: "The most consistently observed differences between children with SLI and control children have been for finite verb inflections and copula and auxiliary forms requiring agreement". This finding has influenced much theorizing about SLI and it has spurred a lot of empirical cross-linguistic research targeting verb forms. However, the results are not always the expected ones, but sometimes there are other structures that emerge as critical, as shown in by the quotations about Spanish, Italian and Swedish (1)–(3).

1. Spanish: "As a case in point, extraordinary difficulty with finite verb inflections stands out as a characteristic of SLI in many languages. However, for Spanish, it appears that articles and clitics are at greater risk" (Bedore & Leonard 2005:223)

2. Italian: "However, relative to MLU controls, Italian-speaking children with SLI do not appear to have special difficulty with most grammatical inflections, in contrast to their extraordinary problem with forms such as articles and clitics" (Leonard 1998:96)

3. Swedish: "Our findings on Swedish children with SLI open up the possibility that the especially serious grammatical impairments in children with SLI extend beyond grammatical morphology, contrary to what has earlier been suggested by research on children with SLI" (Hansson & Nettelbladt 1995:595)

In other studies the analyses have not been able to define which structure is the most vulnerable as shown by the quotations on Japanese and French (4)–(5):

4. Japanese: "The results from Japanese did not fit with any of the theoretical accounts of grammatical deficits in SLI" (Tanaka Welty, Watanabe, & Menn 2002)

5. French: "The results indicate that the spontaneous language of French-speaking children with SLI in the preschool age range is characterized primarily by a generalized language impairment and that morphological deficits do not stand out as an area of particular vulnerability, in contrast with the pattern found in English for this age group." (Thordardottir & Namazi 2007:698)

In sum, it would appear that for some languages there is enough evidence to claim that there is one particular structure presenting difficulties, for others the SLI problem seems to be more general. This may partly depend on the methodology used in gathering the evidence, with studies conducted with large groups of children in experimental settings targeting particular structures and overlooking others. The methodology is obviously connected with the theoretical bias.

3. A study on Swedish children with SLI

The aim of this section is to present a reanalysis of data from Swedish children with SLI (Håkansson 1997, 2001; Håkansson & Hansson 2000) and to propose an explanation to the seemingly contradictory one mentioned above. Three aspects contribute to make this study different from the ones discussed above. The first and most crucial difference is that this study uses PT as its theoretical framework. This means that it analyzes the language production of children with SLI bottom-up, as if they were learners, in order to determine which level of the processability hierarchy they are able to produce. Because the children are diagnosed with SLI, I do not expect them to be able to process all levels; and because the hierarchy is implicational, I expect them to stop at a certain level. If all children stop at the same level of processability, this level may be interpreted as a critical feature of SLI. The other two differences between this study and those mentioned above are methodological and stem from the way data is analysed in the PT framework. The criterion for acquisition is based on productive use, and not on percentage correctness. One minimal pair is enough to show that the child differentiates for example singular from plural form of a noun – this is labelled the emergence criterion. Finally, the production from each individual is analysed and the results are given for individual children (To illustrate the point the results will also be given as group means).

3.1 Grammatical structures in Swedish

In Pienemann and Håkansson (1999) Swedish grammar was outlined from the perspective of processing complexity. Here, some of the morphological and syntactic phenomena will be presented, in the order in which they are predicted to emerge in the acquisition of Swedish. The following structures will be analysed: suffixes on verbs, agreement in VP, subject-verb inversion, and the specific subordinate clause word order. The structures can be found on different levels in Table 1 below. For practical reasons, I will use the PT hierarchy that compiles both morphology and syntax in the same table. Table 1 illustrates the five levels of processing procedures and their morphological and syntactic outcome for Swedish.

Table 1. Processing procedures applied to Swedish (after Pienemann & Håkansson 1999)

PROCESSING PROCEDURE	MORPHOLOGY	SYNTAX	Swedish examples
5 Subordinate clause procedure		specific subordinate clause word order – negation in front of finite verb	där är barnet som inte kan gå (‘there is the child who not can walk’)
4 Sentence procedure		subject-verb inversion after topicalized element	sen kom han in (‘then came he in’)
3 Phrasal procedure	VP agreement NP agreement		äter – har ätit (‘eat – have eaten’) bil – röd-a bil-ar (‘car – red cars’)
2 Category procedure	present – past		hoppa – hoppa-de (‘jump – jumped’)
1 Word/lemma	invariant forms	single constituent	‘words, chunks’

The children are expected to produce the structure in the order bottom – up. Level 1, invariant forms, is not analyzed. At level 2, the category procedure, we find suffixes, which for example, mark simple tense (present and past) on verbs. Level 3 involves an exchange of information between elements within the same phrase, and the unification of the diacritic features is visible via agreement morphology. The VP agreement demonstrates an exchange of grammatical information. There is unification of features between auxiliary and main verb to ensure that only one verb is marked for tense. See (7), below:

- (7) [äter]VP [[har]Aux-pres [ätit]V-supine]VP
 ‘eat’ ‘has eaten’

Swedish perfect tense consists of the auxiliary *har* (‘have’) and a main verb in supine form. The supine is a non-finite form of the verb and cannot be used in isolation in main clauses. Since the compound tense involves an exchange of information between two constituents in the phrase, it is predicted that perfect tense will appear later than present and past tense in the development of Swedish.

At the level of interphrasal morphology (level 4), the different grammatical functions of the constituents in the clause are identified and the verb is placed in second position (V2), in front of the negation. Swedish, in contrast to many other Germanic languages, does not have subject-verb agreement markers. Instead, the processing of this level is realized in the subject-verb inversion, which is obligatory

in topicalized declaratives. Exchange of information between phrases is situated higher than exchange within phrases on the processability hierarchy. Thus the subject-verb inversion is predicted to appear later than the perfect tense.

At the top of the processing hierarchy (level 5) we find the subordinate clause procedure, which is an exchange of grammatical information between main clause and subordinate clause. In Swedish, word order in subordinate clauses is different from the word order in main clauses. The V2-rule is only applied in main clauses, not in subordinate clauses. In subordinate clauses the negation is placed before the finite verb. Subordinate clauses can be expected to be processable after inversion in main clauses is applied, since there is a need of exchange of grammatical information between the clauses in order to treat the subordinate clause as a part of the main clause.

Concluding this section, I will summarize the PT predictions for the order of processability of Swedish structures.

- i. Simple tense (level 2) before compound tense (level 3)
- ii. Compound tense (level 3) before subject-verb inversion (level 4)
- iii. Subject-verb inversion (level 4) before subordinate clause word (level 5)

3.2 Material and methods

Twenty Swedish-speaking children participated in the study, 10 of whom are diagnosed with SLI (aged 4;0 – 6;3, mean age 5;1), and 10 children with typical development (TD; aged 3;1 – 3;7, mean age 3;4). They were matched for utterance length. All children were tested with material designed to create obligatory contexts for the targeted structures (simple and compound tense, NP agreement, subject verb inversion, negated subordinate clauses). The entire procedure was recorded. The interviewer used a coding form to transcribe the elicited utterances. In addition to this form, parts of the dialogue were transcribed. As mentioned above, what is counted is production or non-production of grammatical structures at different levels of processing complexity, not correctness. This means, for example, that irregular verbs inflected as regular by the children (e.g. *dricckte* ‘drinked’ instead of *drack* ‘drank’) are not analyzed as errors but as examples of productive past tense morphology.

4. Results

The results are first given as group means where the children are treated as two homogenous populations. Then their individual performance is discussed, presented in implicational scales. As illustrated in Figure 1, the group means show a difference between the two groups of children.

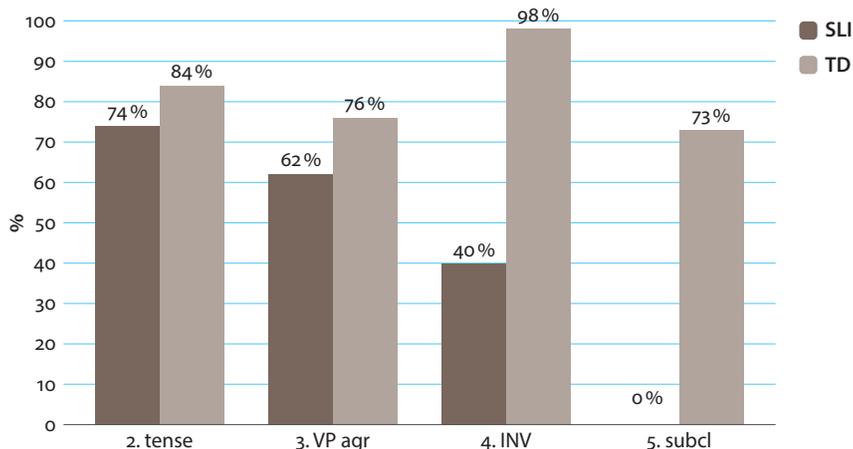


Figure 1. Use of simple tense, VP agreement, subject-verb inversion and subordinate clause word order in % of obligatory contexts by 10 children with SLI, and 10 typically developing children (TD)

At the group level, the results reveal that the TD group outperforms the SLI on all structures: tense, VP agreement, inversion and subordinate clause word order. The differences between the groups are smaller for lexical and phrasal morphology (levels 2 and 3), than for inversion (98% occurrences in obligatory contexts for TD and 40% for SLI) and subordination (40% occurrences for TD and none for SLI). But what about individual variation? Do all children with SLI vary in their production of inversion, or are some children using inversion and some not?

To answer this question and account for the individual variation, the data are now presented in another way, from the point of view of each individual's performance. In the implicational table below the children are ordered according to production or non-production of the structures under discussion. A plus in the table means that there are at least two contrasting structures. For tense, there is systematic form variation on the same verb, and the verb is used both in present and past. For VP agreement the same verb has to be used both in finite form (present or past) and in infinitive form (supine or infinitive) with an auxiliary. For a plus in inversion, there has to be at least two clauses with a topicalized element and subject-verb inversion (plus examples with subject-initial clauses as contrast). Finally, for a plus in subordinate clause word order, there must be at least two cases of negation in front of the finite verb in a subordinate clause and two cases of a negation after the verb in a main clause. The figures indicate how many occurrences of obligatory contexts (e.g. 5/10 means five occurrences in 10 obligatory contexts). Observe that the number of obligatory contexts may differ between children. For example, Fabian produces

seven sentences with topicalized adverbials, whereas Krista only produces three sentences with topicalized adverbials. Neither of them produces the required subject-verb inversion.

Table 2. Implicational order of structures used by ten children with SLI

	Level 2. TENSE		Level 3. VP AGR		Level 4. subj-verb inversion		Level 5. subclause word order	
	±	occ/obl context	±	occ/obl context		occ/obl context		occ/obl context
Fabian	+	5/10	+	2/4	-	0/7	-	0/8
Filip	+	6/10	+	3/5	-	0/9	-	0/9
Greg	+	6/10	+	2/5	-	0/9	-	0/9
Josef	+	5/10	+	3/4	-	0/10	-	0/9
Henrik	+	8/10	+	5/10	-	0/7	-	0/9
Krista	+	9/10	+	3/10	-	0/3	-	0/9
Robert	+	5/10	+	4/10	+	7/10	-	0/0
Hanna	+	10/10	+	6/7	+	3/6	-	0/9
Tony	+	10/10	+	5/10	+	13/13	-	0/9
Hillevi	+	10/10	+	9/10	+	11/11	-	0/10

When the children are ordered this way, it is evident that the variation between individuals is not random, but systematic. The structures seem to have an increasing degree of difficulty. All children are able to produce structures at the lexical and phrasal stages (stages 2–3) but only four children (Robert, Hanna, Tony and Hillevi) are able to produce structures at the inter-phrasal stage (level 4). This result can be interpreted as an indication that the inter-phrasal stage and the next stage are “vulnerable structures” in Swedish.

It is particularly striking that there are no examples of subordinate clause word order in the children with SLI. This structure was elicited by a game, where the children were expected to pick cards from a plate and tell what they had got. For example, “I got the boy who not could swim”. This structure caused major problems for the children with SLI.

5. Discussion

The results of the analyses reveal that the Swedish-speaking children with SLI differ from English-speaking children by not being particularly poor in tense

morphology. They supply tense suffixes to verbs in a consistent manner (74% of obligatory contexts, not very far from the 84% by the younger TD children). Also the analyses of NP and VP agreement show only a small difference between children with SLI and the TD children. The qualitative analysis is revealed in the analysis of inverted word order and subordinate clauses. The inversion rule is used to a much higher degree by the TD children (98%) than by the SLI children (40%). Furthermore, analysis of individual children level reveals that it is not just a matter of lower proficiency in general, but rather that the children are at different developmental levels. Only four of ten children with SLI make use of the subject-verb inversion rule at all. Six children always use subject-verb word order. The specific subordinate clause word order, on the other hand, is a structure that is never used by the children with SLI.

These findings cannot be accounted for by the hypotheses mentioned above: *the feature blindness hypothesis* (Gopnik 1994), *the missing agreement account* (Clahsen 1992), *the extended optional infinitive hypothesis* (Rice & Wexler 1996) nor *the surface hypothesis* (Leonard 1989). All these hypotheses predict a morphological problem, not problems with word order. *The Representational Deficit for Dependent Relationships* (Van der Lely 1998) could possibly be referred to in explaining the results, since both inversion and subordination have to do with linking elements to each other. But this can only be discussed on the basis of mean values, and there is no explanation of the individual differences as shown by the Swedish children with SLI. How can we account for the fact that four of the children use inversion and six of them do not use inversion?

If a developmental perspective is used instead of a view based on deficient representations, it is easier to explain the Swedish results. According to PT, the morphosyntax emerges gradually, in implicationaly ordered stages. The six children who are unable to process and produce subject-verb inversion are at earlier stages in their development than the four children who do use inversion. Thus, the four children who do use inversion are able to produce structures from earlier stages.

Can PT also account for vulnerable structures in other languages? I have argued that it is not the overt morphological or syntactic markings that present problems for SLI children, but rather the level of grammatical processability that underlies them. Due to typological differences between languages, language impairment presents itself in different structures, but they may have the same underlying processing demands. A possible explanation for the observed difference between English and Swedish children with SLI might be that the English and the Swedish subjects happen to be at different levels of the processability hierarchy. Hypothetically, the English subjects may be at level 1, with no productive morphology, whereas the Swedish subjects are able to process level 2, i.e. tense morphology. The explanation can also lie in the typological differences between

Swedish and English. In Swedish, verb tense is less complex since it is possible to separate tense from finiteness, whereas in English tense markings are inseparable from agreement markers. In terms of processability, this would mean that the Swedish simple tense markers are level 2 markers (lexical morphology) whereas the English markers are level 4 markers (S-procedure).

The developmental perspective and PT hierarchy can be used to explain the Italian data as well. As mentioned above, the vulnerable structures in Italian are articles and object clitics, not bound morphology. Object clitics are placed at a high level (S-procedure) of the Italian hierarchy (Di Biase & Kawaguchi 2002) and they can be expected to be processable at a late stage in language development. Tentatively, also the problems with causative and passive morphology described for Japanese children with SLI (Fukuda & Fukuda 2001) can be described within the developmental framework of PT. Below, the problem areas for English, Italian, Japanese and Swedish children with SLI are summarized in a PT hierarchy of morphosyntax.

Table 3. Processing procedures in English, Italian, Japanese and Swedish. Highlighted areas represent structures that have been reported as problematic in children with SLI in respective language

PROCEDURE	OUTCOME	ENGLISH STRUCTURES	ITALIAN STRUCTURES	JAPANESE STRUCTURES	SWEDISH STRUCTURES
5. SUBCLAUSE PROCEDURE	INTERCLAUSAL INFORMATION	cancel inversion in indirect questions		particles <i>ga/wa</i> distinction in sub. clause & main clause	word order distinction in sub. clause & main clause
4. SENTENCE PROCEDURE	INTERPHRASAL INFORMATION	3rd person singular -s	noncanonical placement of clitic object	noncanonical case marking in passive, caus. and benefactive	subject verb inversion in topicalized clauses
3. PHRASAL PROCEDURE	VP AGREEMENT	AUX + V	AUX + V	Vte-V Vte-PROG	AUX + V
2. CATEGORY PROCEDURE	LEXICAL MORPHEMES	tense marking	tense marking	tense marking	tense marking
1. LEMMA ACCESS	WORDS; FORMULAS	single words; formulas	single words; formulas	single words; formulas	single words; formulas

Table 3 demonstrates that the problems for children with SLI are placed at the corresponding level and above for Italian, Japanese and Swedish. The only exception is English, and, as mentioned above, it can be argued that tense marking in English requires other processing procedures than in the other languages, since it has to be combined with the inter-phrasal processing of 3rd person singular to express present tense.

6. Conclusion

One important finding of this study is the value of studying individuals instead of group means. The variation found in the present study is not random but highly systematic. What was obscured by the group means was that the individual Swedish children with SLI differed from the typically developing children to a certain degree. This suggests that it is fruitful to use a developmental perspective and to study SLI children as individuals on different levels, instead of regarding them as a homogenous population with a common deficit.

The findings from the study also highlight the importance of investigating the processes behind surface structures in cross-linguistic comparisons. Processability Theory is a useful paradigm here. The study offers three convincing examples. First, the analysis reveals that what is traditionally seen as “tense morphology” relies on different processes in English and Swedish. In English, present tense in the third person singular cannot be separated from inter-phrasal agreement. In Swedish, the tense suffix only involves a marking of a diacritic feature of the verb, and is therefore processable at level 2. This means that it is easier to process present tense morphology in Swedish than in English. Secondly, Swedish grammar presents other problems. The SLI-children in this study differ from the TD children in the production of word order in topicalized constructions. For the processing of this structure, there has to be an exchange of grammatical information between constituents, just like in the English third person singular. The third crosslinguistic example is the case of object clitics in Italian. This is also a structure that is found at the inter-phrasal stage.

These three examples show that the assumption that English SLI children, Italian SLI children and Swedish SLI children have different problems is indeed superficial; in fact, these problems have the same source: the exchange of grammatical information between constituents at the inter-phrasal level. In other words, the problem is the same, but it is realized in different structures in the three languages. These findings suggest that it might be fruitful for future cross-linguistic research of SLI to use the predictions from PT to find out what the vulnerable structures are, focusing on the inter-phrasal stage of processability.

References

- Bedore, L. M., & Leonard, L. B. (2005). Verb inflections and noun phrase morphology in the spontaneous speech of Spanish-speaking children with Specific Language Impairment. *Applied Psycholinguistics*, 26, 195–225. doi:10.1017/S0142716405050149
- Bishop, D. (2009). Specific Language Impairment as a language learning disability. *Child Language Teaching and Therapy*, 2, 163–165. doi:10.1177/0265659009105889

- Clahsen, H. (1992). Linguistic perspectives on Specific Language Impairment. *Theorie des Lexicons. Arbeiten des Sonderforschungsbereichs* 282. Düsseldorf: Seminar für Allgemeine Sprachwissenschaft.
- Di Biase, B., & Kawaguchi, S. (2002). Exploring the typological plausibility of Processability Theory: Language development in Italian second language and Japanese second language. *Second Language Research*, 18, 272–300.
- Fletcher, P., Leonard, L., Stokes, S. F., & Wonga, A. (2009). Morphosyntactic deficits in Cantonese-speaking children with specific language impairment. In S.-P. Law, B.S. Weekes, & A.W.-Y. Wong (Eds.), *Language disorders in speakers of Chinese* (pp. 75–88). Bristol: Multilingual Matters.
- Fukuda, S., & Fukuda, S. (2001). The acquisition of complex predicates in Japanese specifically language-impaired and normally developing children. *Brain & Language*, 77, 305–320. doi:10.1006/brln.2000.2404
- Gopnik, M. (1994). Impairments of tense in a familial language disorder. *Journal of Neurolinguistics*, 8, 109–133. doi:10.1016/0911-6044(94)90020-5
- Håkansson, G. (1997). Language impairment from a processing perspective. *Working Papers*, 46, 131–151. Lund: Dept of Linguistics, Lund University.
- Håkansson, G. (2001). Tense morphology and verb-second in Swedish L1 children, L2 children and children with SLI. *Bilingualism: Language and Cognition*, 4, 85–99. doi:10.1017/S1366728901000141
- Håkansson, G., & Hansson, K. (2000). Comprehension and production of relative clauses: A comparison between Swedish impaired and unimpaired children. *Journal of Child Language*, 27, 313–333. doi:10.1017/S0305000900004128
- Hansson, K., & Nettelblatt, U. (1995). Grammatical characteristics of Swedish children with SLI. *Journal of Speech, and Hearing Research*, 38, 589–598. doi:10.1044/jshr.3803.589
- Hansson, K., Nettelblatt, U., & Leonard, L. (2000). Specific Language Impairment in Swedish: The status of verb morphology and word order. *Journal of Speech, Language, and Hearing Research*, 43, 848–864. doi:10.1044/jslhr.4304.848
- Hsieh, L., Leonard, L., & Swanson, L. (1999). Some differences between English plural noun inflections and third singular verb inflections in the input; the contribution of frequency, sentence position and duration. *Journal of Child Language*, 26, 531–543. doi:10.1017/S030500099900392X
- Ito, T., Fukuda, S., & Fukuda, S. (2009). Differences between grammatical and lexical development in Japanese specific language impairment: A case study. *Poznán Studies in Contemporary Linguistics*, 45, 221–221.
- Jakubowitz, C., Nash, L., Rigaut, & Gerard, C. (1998). Determiners and clitic pronouns in French-speaking children with SLI. *Language Acquisition*, 7, 113–160. doi:10.1207/s15327817la0702-4_3
- Jørgensen, N. 1976. *Meningsbyggnaden i talad svenska*. Lund: Studentlitteratur.
- Josefsson, G. (2003). Input and output. Sentence patterns in child and adult Swedish. In G. Josefsson, C. Platzack, & G. Håkansson (Eds.), *The acquisition of Swedish grammar* (pp. 95–133). Amsterdam: John Benjamins. doi:10.1075/lald.33.05jos
- Leonard, L. (1989). Language learnability and specific language impairment in children. *Applied Psycholinguistics*, 10, 179–203. doi:10.1017/S0142716400008511
- Leonard, L. (1998). *Children with Specific Language Impairment*. Cambridge, MA: The MIT Press.
- Leonard, L. (2009). Some reflections on the study of children with specific language impairment. *Child Language Teaching and Therapy*, 25(2), 169–171. doi:10.1177/0265659009105891

- Pienemann, M. (1998). *Language processing and second language development: Processability Theory*. Amsterdam: John Benjamins. doi:10.1075/sibil.15
- Pienemann, M. (Ed.). (2005). *Cross-linguistic aspects of Processability Theory*. Amsterdam: John Benjamins. doi:10.1075/sibil.30
- Pienemann, M., & Håkansson, G. (1999). A unified approach towards the development of Swedish as L2: A processability account. *Studies in Second Language Acquisition*, 21, 383–420. doi:10.1017/S0272263199003022
- Rice, M., & Wexler, K. (1996). Towards tense as a clinical marker of Specific Language Impairment in English-speaking children. *Journal of Speech and Hearing Research*, 39, 1239–1257. doi:10.1044/jshr.3906.1239
- Tanaka Welty, Y., Watanabe, J., & Menn, L. (2002). Language production in SLI Japanese pre-schoolers. In E. Fava (Ed.), *Clinical linguistics. Theory and applications in speech pathology and therapy* (pp. 175–195). Amsterdam: John Benjamins. doi:10.1075/cilt.227.14tan
- Thordardottir, E., & Namazi, M. (2007). Specific Language Impairment in French-speaking children: Beyond grammatical morphology. *Journal of Speech, Language, and Hearing Research*, 50, 698–715. doi:10.1044/1092-4388(2007/049)
- Van der Lely, H. (1998). SLI in children: Movement, economy and deficits in the computational-syntactic system. *Language Acquisition*, 7, 161–192. doi:10.1207/s15327817la0702-4_4

Testing the Developmentally Moderated Transfer Hypothesis

The initial state and the role of the L2 in L3 acquisition*

Manfred Pienemann, Anke Lenzing & Jörg-U. Keßler

Paderborn University and Newcastle University, UK / Paderborn University / Ludwigsburg University of Education

This paper focuses on one specific aspect of the Developmentally Moderated Transfer Hypothesis (Pienemann et al. 2005), namely the role of the L2 in L3 acquisition. The research presented in this paper was prompted by the L2 transfer hypothesis put forward by Bohnacker (2006) and Bardel and Falk (2007). According to this hypothesis, learners transfer features from the L2 to the L3, but not from the L1 to the L3. This proposal is partly in conflict with the Developmentally Moderated Transfer Hypothesis which predicts that learners transfer features from the L1 or the L2 to the new language when they are developmentally ready to acquire the features to be transferred, but not before.

The articles by Bohnacker (2006) and Bardel and Falk (2007) are attempted rebuttals of Håkansson et al.'s (2002) work on L1 transfer and aspects of the underlying theory: Processability Theory (Pienemann 1998). The article by Håkansson, et al. presented empirical evidence showing that Swedish learners of L2 German do not transfer V2 at the initial state although both are V2 languages. Bohnacker (2006) and Bardel and Falk (2007) claim that the non-transfer of V2 is due to the influence of the L2. They further claim to have shown in their own study that the initial L3 word order is determined by the L2, irrespective of the structure of the L1 and independently from constraints on processability.

* The authors would like to thank Gisela Håkansson (Lund University) and Bruno Di Biase (Western Sydney University) for their useful comments on this paper. An earlier version of this paper was published in 2013 (Pienemann, M., Keßler, J.-U., & Lenzing, A. (2013). Developmentally Moderated Transfer and the role of L2 in L3 acquisition. In A. Flyman Mattson & C. Norrby (Eds.), *Language acquisition and use in multilingual contexts* (pp. 142–159). Lund University: Travaux de l'Institut de Linguistique de Lund.)

In their response to Bohnacker (2006), Pienemann and Håkansson (2007) demonstrated that Bohnacker's informants had reached an advanced level of acquisition and that this set of data was not suitable to test hypotheses about transfer in the initial state.

In this paper we review the study by Bardel and Falk (2007) and present the gist of an extensive replication of this study. We show that Bardel and Falk's study is based on a very limited database and on theoretical concepts that lack validity, in particular the notion of a 'strongest L2' which is crucial to Bardel and Falk's approach.

Our replication study shows that the initial L3 word order and the initial position of negation is neither determined by the L1 nor by the L2 and that it can be predicted on the basis of processability.

1. Developmental Moderation of Transfer and L2 Transfer in L3 Acquisition

1.1 The Developmentally Moderated Transfer Hypothesis

The Developmentally Moderated Transfer Hypothesis (DMTH) is a component of Processability Theory (Pienemann 1998); it was spelt out in detail with empirical support in Pienemann et al. (2005). The basic idea behind the DMTH is the following: given the architecture of human language processing, the L2 formulator relies on L2-specific lexical information that is essential for grammatical processing. The learner has no *a priori* knowledge of L2-specific lexical information such as the diacritic features of the L2 lexical categories. Therefore, full transfer of the L1 at the initial state would lead to very unwieldy hypotheses. Instead, it is assumed that the L2 lexicon is annotated gradually and that this together with the development of L2 processing procedures permits the learner to build up the L2 in stages. As illustrated in Figure 1, features of the L1 will be able to be utilised

L1 transfer is developmentally moderated:
 "One can only transfer what can be processed"
 =L1 transfer may occur when the given structure can be processed, not before.

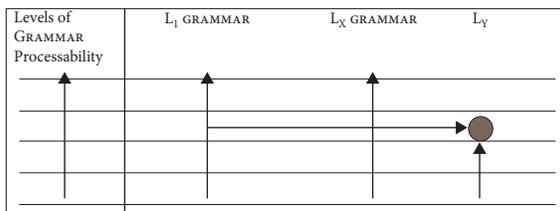


Figure 1. The Developmentally Moderated Transfer Hypothesis (taken from Pienemann 2011: 76)

once the developing L2 system can process them. For this reason, all learners are predicted to follow the same developmental trajectory irrespective of the L1, and positive and negative effects of the L1 will be visible at predictable points of development. In other words, the DMTH does not rule out transfer altogether. Instead, it assumes a selective role of transfer in SLA.

1.2 Håkansson Pienemann & Sayehli (2002)

The study by Håkansson, Pienemann, and Sayehli (2002) provides empirical support for the DMTH. The study focuses on the acquisition of German by Swedish school children. The L1 and the L2 are typologically close and share the following word order regularities in affirmative main clauses:

SVO
 adverb fronting (ADV)
 V2 (verb-second) after ADV.

The following examples of V2 in German and Swedish illustrate the word order similarity in the two languages:

German = V2 Dann kauft das Kind die Banane
 Swedish = V2 Sen köper barnet bananen
 (*Then buys the child the banana*)

Note that in German and Swedish, sentences without V2 are ungrammatical – as shown in the following example:

*Dann das Kind kauft die Banane
 (*Then the child buys the banana*)

Figure 2 gives an overview of the acquisition of key word order patterns in the three Germanic languages that are relevant in the context of this chapter. These developmental patterns are displayed in relation to the corresponding PT levels. The results of this study are summarized in Table 1 below, which treats all learner samples as parts of a cross-sectional study. Therefore, Table 1 represents an implicational analysis of the data which demonstrates that the learners follow the sequence (1) SVO, (2) ADV and (3) INV. In other words, ADV and INV are not transferred from the L1 at the initial state although these rules are contained in the L1 and the L2. This implies that for a period of time the learners produce the following constituent order

*adverb+ S + V + O

which is ungrammatical in the L1 as well as in the L2.

<i>PT level</i>	<i>ESL syntax</i>	<i>Swed. L2 syntax</i>	<i>GSL syntax (Meisel et al. 1981)</i>
6	Cancel INV	–	V-Final
5	Do2nd, Aux2nd	V2	V2
4	Y/N inv, copula inv	–	V-Front
3	ADV-1st, WH-1st, Do-1st	ADV-1st, WH-1st	ADV-1st, WH-1st
2	SVO	SVO	SVO
1	invariant forms	invariant forms	invariant forms

Figure 2. L2 syntactic development in three Germanic languages (selected structures)

This finding is consistent with the DMTH because the structures which are identical in the two languages are not transferred at the initial state. Under the transfer assumption, one would have expected to find all obligatory structures to be present in all samples, particularly V2. However, 10 of the 20 samples consistently violate the V2 rule (i.e. *adverb+ S + V + O) despite the marked ungrammaticality of the resulting structure.

1.3 Bohnacker (2006) and Pienemann & Håkansson's (2007) reply

Bohnacker claims that the late acquisition of V2 in Håkansson's study is due to transfer from English, the L2 of all learners in the sample. She further claims that Swedes learning German as the first L2 will start with V2 because they will transfer this structure from the L1. In other words, Bohnacker assumes full transfer from the L1 to the L2 and from the L2 to the L3, if there is an L3. To support her claims, she carried out a replication of Håkansson's et al. (2002) study. Bohnacker's study is based on a group of six elderly Swedes, half of whom report never to have learned English or German. These informants learnt German mostly in order to be able to communicate with their German-speaking grandchildren. The other three learners had English as their L2 and learnt German as L3. Bohnacker found quantitative differences between the two groups of learners. The group without L2 English showed a higher accuracy in the use of V2 in German. However, Pienemann and Håkansson (2007) demonstrated that all learners in Bohnacker's study had already acquired V2, and the data were not suitable to make any statements about transfer at the initial state. It may be useful to summarise Pienemann and Håkansson's (2007) review to reconstruct the debate.

Table 1. German as L2 by Swedish learners. Implicational scale based on all learners in the study by Håkansson et al. (2002) (taken from Håkansson et al. (2002: 258))

Name	SVO	ADV	INV
Gelika (year 1)	+	-	-
Emily (year 1)	+	-	-
Robin (year 1)	+	-	-
Kennet (year 1)	+	-	-
Mats (year 1)	+	-	-
Camilla (year 2)	+	-	-
Johann (year 1)	+	+	-
Cecilia (year 1)	+	+	-
Eduard (year 1)	+	+	-
Anna (year 1)	+	+	-
Sandra (year 1)	+	+	-
Erika (year 1)	+	+	-
Mateaus (year 2)	+	+	-
Karolin (year 2)	+	+	-
Ceci (year 2)	+	+	-
Peter (year 2)	+	+	-
Johan (year 2)	+	+	+
Sandra (year 2)	+	+	+
Zofie (year 2)	+	+	+
Caro (year 2)	+	+	+

Pienemann and Håkansson (2007) subjected Bohnacker's data to a re-analysis based on the statistics provided in her paper. The re-analysis was necessary because Bohnacker contrasts her claim with Håkansson's claim that Swedes learning German as L2 start with canonical word order. Therefore, Bohnacker's analysis needs to be based on the same approach to data analysis and the same acquisition criteria (i.e. implicational scaling and the emergence criterion). As mentioned above, Bohnacker's own analysis focuses on quantitative differences between learners. Pienemann and Håkansson (2007) re-analysed Bohnacker's data in form of an implicational analysis using the emergence criterion. The re-analysis is presented in Table 2.

Table 2. Re-analysis of Bohnacker's sample using the same criteria as in Håkansson et al. (2002)

L2 or L3?	Informant	SVX	ADV	SEP (%)	V2 (%)
L2 German	Marta 1	+	+	12 ¹	100
L2 German	Marta 2	+	+	12	100
L2 German	Marta 3	+	+	70	100
L2 German	Algot 1	+	+	30	100
L2 German	Algot 2	-	-	-	-
L2 German	Algot 3	+	+	85	95
L2 German	Signe 3	+	+	62	100
L3 German	Rune 1	+	+	8	55
L3 German	Rune 2	+	+	8	44
L3 German	Rune 3	+	+	76	58
L3 German	Gun 1	+	+	45	55
L3 German	Gun 2	-	-	-	-
L3 German	Gun 3	+	+	70	57
L3 German	Ulf 3	+	+	61	52

Table 2 is laid out as follows. The first column states whether German is the L2 or the L3 of the informant. The second column identifies the sample by informant name and 'data point' number. The column entitled 'SVX' states whether the sample contains examples of canonical word order (using the emergence criterion) where '+' means 'acquired'. The column 'ADV' does the same for structures with non-subjects in initial position. The column 'SEP' lists the relative frequency of two verbs (aux + V) appearing in a non-adjacent position (i.e. XVYV). This structure occurs in German, but not in Swedish. The last column lists the relative frequency of V2 application. In other words, the columns from SVX to V2 are arranged in the order of acquisition that has been found in many previous SLA studies and that was initially identified by Meisel, Clahsen, and Pienemann (1981).

It is easy to see that all four target structures meet the emergence criterion for all of the informants, no cell of the implicational table is empty (apart from missing data for Algot 2 and Gun 2), no learner slides back, and thus the scalability of Table 2 is 100%. This means that all structures under discussion, including V2, had

1. The figures for Marta 1 + 2 and Rune 1 + 2 are presented as averages of the two sessions by Bohnacker.

already been acquired at the first point of data collection. In other words, all informants had acquired V2 (and all the other relevant structures) at the beginning of the study. This is the strongest reason why the study is not suitable to test the initial word order of Swedish first-time learners of German. Given that the learners had already acquired all the structures under investigation at the beginning of the study, including V2, they are simply too advanced to make any statement about the INITIAL state of their interlanguages.

One might object to this conclusion about the level of acquisition of the six learners in Bohnacker's corpus on logical grounds, because full transfer from Swedish would always imply that all structures contained in Table 2 need to be present from the start. However, this hypothetical possibility would apply only to one subgroup of Bohnacker's sample: the learners with L1 Swedish and L2 German. For the other subgroup with L1 Swedish, L2 English and L3 German she predicted transfer from L2 to L3. Given that V2 is not part of English, these learners should not acquire V2 at the initial state. However, Table 2 shows that all learners from this group also display clear evidence of V2 in the first interview. Therefore, the full transfer assumption is not compatible with the evidence presented in Bohnacker's study.

Nevertheless, there is one striking difference between the L2 and the L3 group. The learners without exposure to English display a native level of performance for V2, whereas learners with previous exposure to English do not. This is highly compatible with the DMTH, which predicts that transfer will not appear before the structure to be transferred can be processed by the interlanguage system. However, when structures from the L1 or L2 are processable, they may be transferred to the target language, and this may lead to differential patterns of language use in groups of learners with different L1s (or L2s). We ascertained above that all informants in Bohnacker's corpus have reached the acquisition level marked by 'INV'. Therefore, V2 is readily processable by all learners in this corpus, and the group without knowledge of English can make recourse only to their knowledge of V2 that can be transferred at this point of development, whereas the group with English as the first L2 can transfer two competing rules that match the structural condition for V2, i.e. either XVSY or XSVY. Therefore, the given learning condition facilitates the accuracy with which the L2 group uses V2 compared with the L3 group.

1.4 Bardel & Falk (2007)

Bardel and Falk (2007) also carried out a study that was designed to refute the DMTH. It may be useful to first consider Bardel and Falk's critique of the study by Håkansson et al. (2002). We will then review Bardel and Falk's (2007) own study.

Håkansson et al. (2002) argue as follows about the role of L2 transfer in their study. If the lack of V2 transfer from Swedish to German in their data was due to the transfer of SVO from L2 English, then why were other features of English not transferred, such as ‘adverb-first’ (ADV) or ‘particle’ (e.g. ‘put it on’)? Bardel and Falk respond to this as follows: Adverbs in initial position are missing only in a few of the learners in Håkansson et al.’s study, and adverbs are optional anyway. So there could be any number of reasons for their absence. They conclude that “[...] the absence of a part of speech in oral production can hardly be taken as an argument against transfer.” (Bardel & Falk 2007: 466).

In our view, Bardel and Falk’s line of reasoning ignores the context in which the analysis of word order constellations was conducted in Håkansson et al.’s study where the following word order rules were studied as part of a distributional analysis of the ILs of all 20 informants: SVO, ADV, V2. The corresponding table from Håkansson et al. (2002) with the full distributional analysis was repeated above as Table 1.

As Håkansson et al. (2002) pointed out, Table 1 shows a very strong implicational relationship of the three rules (SVO, ADV and V2) with a scalability of 100%. In other words, it shows the following strict implicational relationship:

$$\text{SVO} > \text{ADV} > \text{V2}$$

This is supported by a comparison of the development of those learners who were recorded at two points in time within a one-year interval.

In light of these circumstances, it is unlikely that for the learners who display the feature SVO only the absence of ADV is a pure coincidence, because as part of their IL grammar that follows canonical word order, they systematically place adverbs and adverbials in final position as shown in the following example:

- (1) Der Mann gehen nach Hause.

In other words, the fact that adverbs and adverbials are not fronted in a subset of the data is not merely a reflection of the presence or absence of these structures or lexical category in the sample. Instead, the learners in this group systematically placed adverbs and adverbials in final position, whereas the learners with ADV alternated between initial and final position depending on pragmatic conditions.

1.5 Bardel and Falk’s study

Bardel and Falk (B&F) present an empirical study with two distinct sets of data to support their L2 transfer hypothesis. Data set A is based on five learners of Swedish as an L3 who were exposed to a sequence of ten 45-minute lessons of Swedish. The typology of the learners’ languages can be summarised as follows:

# learners	L1	L2	L3
3	+V2	-V2	+V2
2	-V2	+V2	+V2

This design permits the specific effect of the L2 to be tested empirically. Data set B was set up in a similar manner with respect to the typology of the learners' languages, only this time the learners ($n = 4$) were recorded once in one-off individual one-to-one lessons lasting about 45 minutes. According to B&F (2007:470), the learners were "absolute beginners" and had no prior knowledge of Swedish.

The typology of the learner's languages for 'Data set B' is summarized in Table 3.

Table 3. The learners and their knowledge of V2 languages, data collection B (taken from Bardel and Falk 2007:472)

Learner ²	Sex	First language	Second language	Target language
EN4	F	Swedish +V2	English	Dutch +V2
EN5	M	Swedish +V2	English	Dutch +V2
D/G3	M	Italian	German/Dutch +V2	Swedish +V2
D/G4	M	Albanian	German +V2	Dutch +V2

The study focused on the acquisition of negation. Using example sentences, B&F provide an implicit distribution of the position of the negator for the languages that are relevant in this study (with a focus on sentence negation in declaratives) which can be summarized as follows:

S V neg	Swedish, Dutch, German
S cop neg A	Swedish, Dutch, German, English
S Aux neg V	Swedish, Dutch, German, English
S DO neg V	English
S neg V	Italian, Hungarian, Albanian
S neg cop A	Italian, Hungarian, Albanian
S neg aux V	Italian, Hungarian, Albanian

B&F differentiate between two types of languages in this list, namely (1) languages with preverbal negation (Italian, Hungarian and Albanian) and (2) languages

2. The code EN is used for the learners who speak English as an L2, and the code D/G refers to those learners whose L2 is Dutch and/or German.

with postverbal negation (Swedish, Dutch, German, English). In addition, B&F (2007: 469) make the following assumption about negation in English – following Chomsky (1986):

Verb raising in English (which is not a V2 language) distinguishes thematic from non-thematic verbs, and this has a bearing on the surface pattern of the English negative clause. While non-thematic verbs raise to IP and leave negation in a post-verbal position, thematic verbs remain, uninflected, in the VP [...]

B&F contrast this analysis with that of other Germanic languages in which, according to their analysis, the position of the negator results from V2, which in turn does not differentiate between thematic and non-thematic verbs. Bardel and Falk capitalise on this difference between English and the other Germanic languages in testing their L2 transfer hypothesis. On this basis, they predict that according to the L2-transfer hypothesis “[...] L2 speakers of Dutch/ German [...] will place negation post-verbally as in Swedish, while the other group who have English as an L2 [...] will distinguish between thematic and non-thematic verbs in relation to negation placement, since this is a property of English.” (Bardel & Falk 2007: 474).

Unfortunately, data set A yielded a very small quantity of data. Given that the study focuses on transfer at the initial state the first recoding is of special significance. However, this recording merely contains an average of less than two sentences per learner and structure for the group with a V2-L2 and an average of just over one sentence per learner and structure for the group with a non-V2-L2. For the other recordings the data quantity was even smaller (cf. Bardel & Falk 2007: 475).

This very small amount of data is insufficient for any standard analysis of lexical or syntactic variation aimed at excluding the use of formulae. At the same time, both groups of learners produce examples of pre-verbal negation and post-verbal negation, although the L2 transfer hypothesis predicts a different distribution. Therefore, the relevance of this set of data for the issue of L2 transfer remains to be demonstrated.

Data set B consists of an average of about seven relevant sentences per learner and structure, and the distributional analysis for the four learners shows that the Dutch/German group does not produce pre-verbal negation. In contrast, the English group produces both pre-verbal and post-verbal negation – depending on the presence or absence of non-thematic verbs.

At first glance, this observation may be judged as support of the L2 transfer hypothesis. However, there are two problems with B&F’s study in data set B: (1) The exact status of the learners’ L2 has not been identified and (2) the role of repetitions and chunks in very early formal L2 learning has not been considered.

Obviously, identifying the exact status of the learners' L2 is vital in the context of testing a hypothesis that assigns a special status to transfer from the learner's L2. The learners in data collection B and their knowledge of V2 languages are reported by B&F as shown in Table 3. For data collection A, B&F inform the reader that they asked the learners to self-rate the proficiency of their L2s, and thereby B&F identify the learners' 'strongest L2' which is then recorded under 'second language' in the language profile of the learners for data collection B, this procedure is not mentioned explicitly. One can only assume that it was the same for both sets of data and that the L2s shown in Table 3 are the strongest L2s of the learners. B&F do not mention the other L2s of the learners explicitly. However, the language policies of their countries of origin and their study/work situation suggest very strongly that they also speak other L2s – like the informants in data collection A. B&F state that learner D/G3 “[...]was found via the European Parliament [...]” (Bardel & Falk 2007: 472) and learner D/G4 “[...] was found via the University of Stockholm [...]” (Bardel & Falk 2007: 472). One can assume with near-certainty that the Italian learner recruited through the European Parliament working in Brussels has English as one of his L2s, because English has been prevalent among the language subjects in Italian schools for well over two decades, and English is also the lingua franca at and around the European Parliament.

If the two learners of the V2-L2 group (D/G3 and D/G4) also have a non-V2-L2 – as it appears to be the case – the results of the distributional analysis of data set B appear to be far less straight-forward to interpret than it seemed at first. The absence of pre-verbal negation would then have to be due exclusively to the effect of the strongest L2 which would need to override possible effects of other L2s that do have pre-verbal negation. In fact, the same line of argument would apply to data set A as well because it also contains other L2s besides the 'strongest L2'.

In fact, this line of argument would be required as a matter of principle for B&F's L2-transfer hypothesis to be internally consistent. If one could attribute differential effects to just any L2 in a post-factual manner, the explanatory power of the L2 transfer hypothesis would be eroded. Alternatively, one would need to face a much more challenging task: to design a testable hypothesis of partial L2 transfer that also includes the effects of additional and typologically different L2s.

There are two issues that follow from the 'strongest L2-assumption': (1) how does one measure and define the strongest L2 and (2) what is the theoretical motivation for it? Referring to the first issue, B&F admit that self-rating – which they relied on – “[...] may not be an objective method of identifying exact proficiency in a language, but it would not have been feasible to test proficiency level in all background languages in a precise way.” (Bardel & Falk 2007: 471). We would

like to add that proficiency may not even be the right concept that captures the notion of ‘strongest L2’. This brings us straight to issue (2), the theoretical basis of the notion ‘strongest L2’. The way the term is used by B&F is reminiscent of the notion ‘dominant language’ in the context of research on bilingualism. In a review article on measuring bilingualism Pienemann and Keßler (2007) show that a multitude of different approaches to capturing language dominance has been discussed in the past five decades without an operational consensus. Obviously, B&F do not provide an explicit rationale for what a strong L2 is and why it should have a privileged status in non-native language acquisition. There may be a hint at what they have in mind in the very last sentence of their article: “[...] in L3 acquisition, the L2 acts like a filter, making the L1 inaccessible.” (Bardel & Falk 2007: 480). This begs the question what it filters, besides the L1 – the weaker L2s? All of them and all features? And where does this happen? In production and comprehension? In the Formulator (e.g. Levelt 1989), in the bilingual Formulator (de Bot 1992), in the lexicon? And how is this performance-based filter related to linguistic knowledge in the various languages? How can all of this be represented? – Obviously, B&F’s hypothesis is not embedded in any explicit theoretical approach, linguistic or psycholinguistic, and therefore cannot be operationalised.

The second point that needs to be considered for ‘Data set B’ in B&F’s study is the status of formulae. This is relevant here because the data were collected in one single session without any previous contact with the target language. In very early L2 classes, learners’ utterances often consist of formulae and repetitions of the teacher’s utterances, and the structures these appear to contain are not generated by their newly developing non-native formulator. Instead, they are unanalysed large entries in the lexicon. Therefore special care needs to be taken to distinguish between formulae/repetitions and productive learner utterances. Pienemann (1998) showed that a mere count of the occurrence of structures in an L2 corpus does not reveal the underlying learner system and can be rather deceiving. He argued that what is required instead is a test of the null-hypothesis for every structural context. For instance, is a morpheme that marks plural in the target language used in plural contexts only in the learner language or also in non-plural contexts? If it is used in both contexts with a similar frequency, it is obviously not a productive part of the IL grammar. B&F did not test the null hypothesis.³ Therefore, we cannot rule out that the apparent distribution in of pre-verbal and post-verbal negation is based on formulae or repetitions.

3. In fact, data set A would be far too small for this purpose anyway.

2. The PALU⁴ study: Minimal exposure to the L2

Our reservations about the B&F study prompted us to replicate key aspects of that study – with a view on overcoming the shortcomings of their study. One key aspect of the design of our replication study was to differentiate between formulaic echoes of teacher utterances and creative L2 productions.

In the PALU study, the L3 is Swedish, and the L1 is German. The informants have different L2s. In keeping with the DMTH outlined above, it is hypothesised that all learners follow a strictly implicational sequence of developmental stages and do not transfer any structures from their first or second language before they are developmentally ready. Also, we expect all structures they produce to be in line with the processability hierarchy for Swedish as an L2 outlined in Pienemann (1998: 190 ff.). However, learners are expected to be able to repeat phrases and sentences and to use structures as fixed formulae.

2.1 Research design

The data collection for the PALU study was conducted at the University of Paderborn, Germany. The participants were seven German students of linguistics all of whom were fluent speakers of English with high C-test scores (cf. Grotjahn 1992). Three of the students had some prior knowledge of Swedish: C01 and C02 attended a one-semester Swedish course and C01, C02 and C05 took part in a comparative course of Nordic languages. The other four students had no prior knowledge of the Swedish language and its structure. However, as all seven informants were students of linguistics, and as the curriculum includes courses in both theoretical and comparative linguistics, they all had some meta-linguistic awareness.

The participants can be divided into two groups according to their knowledge of verb-second languages other than German (cf. Table 4 and 5). The first group (group A) consisted of four learners with English as their (first) L2 who learned one or more Romance languages afterwards (e.g. French, Italian). The second group (group B) comprised three learners who also had English as their (first) L2 but who additionally learned a V2 language (e.g. Dutch).

4. PALU refers to the universities of Paderborn and Ludwigsbu

Table 4. The informants and their knowledge of languages, Group A (no V2 languages)

Learner	Sex	First language	Additional Languages
C03	F	German	English French Latin Arabic
C04	F	German	English Latin French Spanish Russian
C05	F	German	English French Italian Chinese
C07	F	German	English French Spanish Portuguese Italian

Table 5. The informants and their knowledge of languages, Group B (+ V2 languages)

Learner	Sex	First language	Additional Languages
C01	F	German	English French Spanish Italian Portuguese Turkish Dutch + V2 Swedish + V2
C02	F	German/Russian	English French Spanish Italian Swedish + V2
C06	F	German	English Latin Italian French Dutch + V2

In order to test the hypotheses outlined above, the framework for data collection consisted of three main components: (1) a lesson in Swedish, (2) a session with four communicative tasks, which took place after the lesson and (3) a post-test that was conducted two weeks after the Swedish lesson.

Prior to the Swedish lesson, the informants listened repeatedly to a recording of forty Swedish words (nouns, verbs, adjectives, adverbs) that were related to the communicative tasks used in the subsequent lesson while looking at picture cards that illustrate these words. This was done to ensure that the students familiarised themselves with the vocabulary of the lesson and the related tasks. After this exercise all seven informants participated in a 30 minute 'one-to-one' lesson which was conducted in Swedish by a native speaker who is a university lecturer of Swedish.

During the lesson, a dialogue was rehearsed and a number of daily activities were described. The vocabulary introduced in the lesson was mainly based on the recorded words. The input provided by the teacher consisted of structures that were located at the different stages of the PT hierarchy including different forms of negation and the occurrence of adverbs in varying positions in the sentence.

The overall aim of the lesson was to provide large numbers of contexts for the students to repeat utterances and thus to provide an environment for the production of formulaic speech (cf. Pienemann 2002; Aguado 2002). This focus on formulaic speech permits us to test our hypothesis that learners of a foreign language are able to repeat advanced L2 structures which they are unable to produce productively.

The communicative tasks were structured in a way to ensure that they will elicit sentences that are different from the material that was rehearsed in the lesson. This precaution was taken to ensure that creative L2 constructions produced by the learners are not copies of rote-memorised sentences. The post-test followed the same format as the session with communicative tasks.

2.2 Results

The results of the PALU study for V2 are presented in Table 4, which is laid out as follows. The first column lists the informants; the second marks the presence of SVO. The third column details the frequency of the structure '*advSVO' which is ungrammatical in the source language and the target language. The column headed 'V2' lists information about the presence (+) or absence (-) of V2 in the sample of the individual learners. The column 'L2=V2' specifies if the informant acquired a V2 language as an L2 before the study. The next column specifies if the informant has learnt Swedish before, and the last column gives the frequency of V2 imitations in each sample.

Table 6. Swedish word order in the PALU study

Informant	SVO	*adv SVO	V2	L2 = V2?	Swedish before?	Imitation of V2
C03	+	14	-	-	-	16
C05	+	25	-	-	-	14
C07	+	-	-	-	-	10
C04	+	-	-	-	-	20
C01	+	30	-	+	+	30
C02	+	15	-	+	+	15
C06	+	13	-	+	-	9

As can be seen from Table 6, all learners produced novel SVO sentences even though the 'lesson' consisted only of sentence repetitions. This shows that the input was sufficient to stimulate language production after minimal input. As can be seen from the last column, all learners were also able to repeat V2 sentences correctly even without any previous input in Swedish. This observation confirms our hypothesis that learners are able to store and repeat sentences containing advanced structures. In contrast, none of the learners produced V2 structures spontaneously⁵ (column 4). As can be seen from the data shown in column 3, five of the seven learners did produce fronted adverbs and adverbials. In other words, these learners did not produce V2 although they produced the structural condition for V2. The reader will recall that adverb fronting is a structural condition for V2 in the informants' L1 as well as in the target language, i.e. in a situation where transfer at the initial state would have been expected under the full transfer assumption. As can be seen in column 5, non-production of V2 after adverb fronting appears with learners who acquired non-V2 languages before Swedish as well as with learners who acquired V2 languages as second languages and even with the two informants who had learned Swedish before.

In other words, our corpus does not contain a single example of V2 even though the learners know this structure from their L1 and they had plenty of

5. We use a minimal definition of 'spontaneous production' in this context. For the purpose of this study we assume that structures which are not copies of the previous utterance are produced spontaneously. This minimal definition has to be seen in the context of the hypothesis we tested, namely that at the initial state advanced structures such as V2 can be repeated straight after a stimulus sentence has been presented, but that learners will not be able to produce this structure spontaneously. This minimal definition of 'spontaneous production' ensures that our hypothesis is highly falsifiable. It ensures that 'unwanted' data cannot be classified as formulaic copies.

opportunity to use it. This finding constitutes strong evidence supporting the DTMH.

Apart from the focus on V2 the data were also analysed in relation to the position of the negator. For sentence negation the position of the negator in declarative sentences with lexical verbs distributes as follows in the three Germanic languages relevant in this study:

German	V+neg
English	Do+neg+V
Swedish	V+neg

In studies of second language acquisition, forms such as ‘don’t’ are often treated as one lexical entry which serves as a negator in the interlanguage. This approach is useful in early SLA when the corpus does not contain any lexical or morphological variation of the negator or the auxiliary. However, B&F hypothesize that L3 learners are able to transfer developmentally advanced structures from the L2. Therefore our analysis treats the negator and the verbal element ‘do’ as two distinct constituents as they appear in the target language.

Table 7 displays the distributional analysis of the use of negation in the sample. The first column identifies the informant. The next three columns contain the counts of examples of the key structures ‘neg+V’, ‘V+neg’, ‘aux+neg+V’ and Do+neg+V. The next two columns specify if the learners have an L2 containing V2 and if they have learned Swedish before.

Table 7. Negation

Informant	neg V		V neg		aux+neg+V	Do+neg+V	V2=L2?	Swedish before?
C03	14	+	1	-	0	0	-	-
C05	17	+	0	-	0	0	-	-
C07	0	-	16	+	0	0	-	-
C04	12	+	4	(-)	0	0	-	-
C01	1	-	16	+	0	0	+	+
C02	15	+	0	-	0	0	+	+
C06	indiv. strategy		indiv. strategy		indiv. strategy	indiv. strategy	+	-

Our analysis focuses on the first six informants in Table 7 because informant C06 produced exclusively lexical forms of the negator that do not exist in Swedish and were not contained in the input. Learners C1 and C3 have previously learned Swedish. Therefore any structures appearing in their sample may be residual effects

of their knowledge of Swedish. Of the four remaining learners three produce preverbal negation, i.e. the developmentally earlier structure. The only exception is C7. This observation supports the DMTH, which predicts that developmentally late structures can only be transferred when the interlanguage is ready for it.

Given that all learners also have English as their L2 and that they all speak it at a very high level of proficiency, the above data also permit a test of the L2 transfer hypothesis. According to this hypothesis the learners would be expected to transfer 'Do-insertion' from English. This was tested in the 4th and the 5th column of Table 7 above. As can be seen in column 4, none of the learners transfers 'Do-insertion'. Column 5 shows that in addition *aux+neg+V*⁶ also does not appear. This finding boldly contradicts B&F's prediction that L3 learners will transfer structures from L2 English.

2.3 Summary and discussion

The starting point of the current debate about the role of the L2 transfer was the study by Hakansson et al. (2002) who found that Swedish learners of German do not transfer V2 although V2 is part of German and Swedish grammar. They explained this phenomenon with reference to PT and the assertion that V2 is too complex to be processable at the initial state.

Bohnacker (2006) hypothesized that the non-transfer of V2 in the study by Hakansson et al. (2002) was due to the effect of English which was the L2 of all learners in the study by Hakansson et al. (2002). Bohnacker (2006) presented her own study of German as L2 which she conducted with elderly Swedish learners of German who were reported not to have any English as L2 as an intervening variable. She showed that these learners were more accurate in the use of V2 than Swedish L2 learners of German who had learned English as an L2. However, the reanalysis of her data in the form of an implicational analysis demonstrates that all learners in her study had already acquired V2. Therefore, her study did not show that V2 is transferred at the initial state and that the previous knowledge of intervening L2 is limited to an effect on grammatical accuracy.

Bardel and Falk (2007) repeat Bohnacker's claim about the effect of L2 English in the acquisition of German by Swedes. These authors present two studies of the acquisition of V2 languages by learners with different L1s and L2s. They aim to demonstrate distinctive L2 effects in L3 acquisition. One of their studies did not contain sufficient data to demonstrate their claim. The other study which was based on four informants who received 45 minutes' exposure relied heavily on the dominance of one of the L2s of the learner. B&F claim that it is the dominant

6. Note that for the purpose of this analysis 'aux' includes auxiliaries and modals.

L2 from which the learner makes full transfer and that this process suppresses L1 knowledge. In this paper we have shown that this position is inconclusive at a conceptual level, because it does not specify how the mind identifies the strongest L2, nor what to do with other L2 knowledge, how this transfer-filter relates to linguistic knowledge and performance nor how it can be operationalised. In addition, it remains unclear in B+F's study which of the limited utterances produced by the learners after a one-off 45 minute session were mere repetitions of input and which were productive utterances.

Our own study was based on the acquisition of Swedish as a target language by seven learners with different L2s. All of the learners were highly proficient in English. All had several other L2's, and three of them had L2s with V2. All learners were able to repeat V2 sentences in Swedish, but none of the learners was able to use this structure productively. This observation strongly supports the DMTH and it also demonstrates that formulae can easily distort research findings in studies focusing on the initial state.

Our study also focused on the acquisition of negation. We were able to show that except for one learner, all relevant informants acquired the developmentally earlier structure ('neg+V'). We also demonstrated that none of the learners transferred 'Do-insertion' from English, nor any similar structure. Both these observations clearly support the DMTH.

The above study focused on the restrictive effects of processability on transfer. In this context, the DMTH might be misconstrued as a non-transfer approach. This would be incorrect. As we pointed out in the first section of this chapter, the DMTH defines constraints on transfer. This implies that restrictive and productive effects of the L1 will materialise at predictable points of development. Pienemann et al. (2005) reported on productive effects of L1 transfer at points of processability. Also, the DMTH does not in any way exclude the possibility that features of the L2 may be transferred to the L3. However, as we have shown above, for any L2 transfer hypothesis to make a genuine contribution to a theory of SLA it would need to be fully operationalised and theoretically motivated. Anything less would be no different than any of the speculative approaches which the field has experienced over the past few decades.

References

- Aguado, K. (2002). Formelhafte Sequenzen und ihre Funktionen für den L2-Erwerb. *Zeitschrift für angewandte Linguistik*, 37, 27–49.
- Bardel, C., & Falk, Y. (2007). The role of the second language in third language acquisition: The case of Germanic syntax. *Second Language Research*, 23(4), 459–84.
doi:10.1177/0267658307080557

- Bohnacker, Ute. (2006). When Swedes begin to learn German: From V2 to V2. *Second Language Research*, 22, 1–44. doi:10.1191/0267658306sr2750a
- Chomsky, N. (1986). *Barriers*. Cambridge, MA: The MIT Press.
- de Bot, K. (1992). A bilingual processing model: Levelt's 'speaking' model adapted. *Applied Linguistics*, 13, 1–24. doi:10.1093/applin/13.1.1
- Grotjahn, R. (Ed.). (1992). *Der C-Test. Theoretische Grundlagen und praktische Anwendungen*. Bochum: AKS-Verlag.
- Håkansson, G., Pienemann, M., & Sayehli, S. (2002). Transfer and typological proximity in the context of second language processing. *Second Language Research*, 18(3), 250–273. doi:10.1191/0267658302sr2060a
- Levelt, W. (1989). *Speaking. From intention to articulation*. Cambridge MA: The MIT Press.
- Meisel, J., Clahsen, H., & Pienemann, M. (1981). On determining developmental sequences in natural second language acquisition. *Studies in Second Language Acquisition*, 3(2), 109–135. doi:10.1017/S0272263100004137
- Pienemann, M., & Keßler, J.-U. (2007). Measuring bilingualism. In P. Auer & L. Wei (Eds.), *Handbook of applied linguistics. Multilingualism. Vol. 5*. Berlin: De Gruyter.
- Pienemann, M. (1998). *Language processing and second language development: Processability Theory*. Amsterdam: John Benjamins. doi:10.1075/sibil.15
- Pienemann, M. (2002). Unanalysierte Einheiten und Sprachverarbeitung im Zweitspracherwerb. *Zeitschrift für angewandte Linguistik*, 37, 3–26.
- Pienemann, M., & Keßler, J.-U. (2007). Measuring bilingualism. In Auer & Wei (Eds.), pp. 247–275.
- Pienemann, M. (2011). L1 transfer. In M. Pienemann, & J.-U. Keßler (Eds.) *Studying Processability Theory* (pp. 75–83). Amsterdam: John Benjamins. doi:10.1075/palart.1.06lit
- Pienemann, M., Di Biase, B., Kawaguchi, S., & Håkansson, G. (2005). Processability, typological distance and L1 transfer. In M. Pienemann (Ed.), *Cross-linguistic aspects of Processability Theory*. Amsterdam: John Benjamins. doi:10.1075/sibil.30
- Pienemann, M., & Håkansson, G. (2007). Full Transfer vs. Developmentally Moderated Transfer. *Second Language Research*, 23(4), 485–494. doi:10.1177/0267658307080332

PART II

Theory Assessment

The ‘tense’ issue

Variable past tense marking by advanced end-state Chinese speakers of L2 English

Yanyin Zhang & Bo Liu

The Australian National University

Chinese learners of L2 English tend to show variable past tense *-ed* marking even at an advanced proficiency level. The source of this problem has been explored and debated extensively but no conclusion has been reached (see Beck 1997; Lardiere 1998a/b; Hawkins & Liszka 2003). In this study we continue the investigation by testing two hypotheses: (a) the variable past tense marking is a reflection of the training learners have received during their university study, and (b) rigorous training discourages the ‘bad choices’ being made. Through examining the L2 English speeches of 9 advanced end-state L1 Chinese speakers who had learned English in either top-notch or non-top-notch programmes in china, we found that rigorous training programmes did indeed lead to a high level of ultimate attainment in the past-marking albeit not at the native-like level. It also inhibits ‘bad choices,’ ensuring a uniform high rate of L2 English morphological marking.

1. Introduction

“Tense has the reputation of being the most tortuous of grammar... Though native speakers of English use its tense system effortlessly, it often bewilders people who learn it as adults.”

(S. Pinker. *The Stuff of Thought*. 2007: 193; 197)

The variable marking of the past tense *-ed* by L2 learners of Chinese background has intrigued SLA scholars for a long time. A number of studies found that even near-native speakers of this L1 group marked past tense on regular verbs variably, either at a rate far below the criterial level¹ (Lardiere 1998a), or below the

1. The criterial level is not a constant. In SLA, it is usually based on accuracy to assess mastery. Different researchers set different criterial targets, usually above average, for example, 60% in Vainikka & Young-Scholten (1994), 70% in Eubank & Grace (1998). Pienemann (1998) proposes the ‘emergence criterion’ to assess language acquisition. For details, see Pienemann (1998) and Pallotti (2007).

rate of suppliance by similar L2 English learners of L1 German and L1 Japanese backgrounds (Hawkins & Liszka 2003). From the perspective of Processability Theory (Pienemann 1998), the past *-ed* sits at Stage 2 of the 6-stage developmental hierarchy for English morphology. While all the Chinese subjects in these studies demonstrated their skill to process this morpheme as measured by the emergence criterion, they fell short of a high level of mastery as measured by the accuracy criterion when compared to the German and Japanese subjects. The question is: what might be the possible reasons or sources for this marking variability?

Beck (1997) carried out a series of experiments to test the L2 inflection-attachment system, i.e., the ability to generate inflectional forms and attach them to the stems of regular verbs. The results showed that Chinese L2 learners of English (at the proficiency level of minimum TOEFL530) did have the morphological knowledge of regular English past tense inflection, and that their L2 competence did not involve inflection-attachment 'deficit' (see also Hawkins & Liszka 2003; Prévost & White 2000). Beck hypothesized that the L2 'impairment' might be in the domain of syntax.

However, Lardiere's (1998a/b) case study of an end-state L2 speaker of English (Patty²) provides evidence contrary to Beck's hypothesis. Although Patty's past tense marking was consistently low (34% overall and 5.80% on regular verbs) (Lardiere 1998a: 16; 2003: 184),³ she did not seem to have problems with syntax as attested by her correct production of English negation, adverb placement, pronominal case marking and a variety of CP clauses. This led Lardiere to conclude that the 'deficit' shown in her past tense marking was domain-specific, i.e., it was confined within morphology, with no connection to her L2 syntax (1998a/b, 2000, also see Eubank & Grace 1998).

If variability in the L2 past tense morphology displayed by advanced L2 English speakers of Chinese background is not due to their morphological knowledge, nor to their underlying syntactic competence, is it possible that the source of the problem comes from L1 phonotactic constraints? Since Mandarin Chinese syllable structure permits only alveolar and velar nasals in the coda position, it is likely that the phonetic realization of the past marker *-ed* [t/d] is compromised in speech production. If that were true, similar effects would be observed in the

2. Unlike Hawkins and Liszka's (2003) informants and the informants in the current study, it is not clear if or how Patty's general English proficiency was assessed. Living, studying and working in an English speaking country for many years is not a guarantee for high English proficiency.

3. Patty's agreement marking on nonpast 3sg thematic main verbs (e.g., *He works everyday*) was even lower, less than 5% in the second and the third recordings (Lardiere, 1998b: 366).

English past participles of regular verbs (*have jumped*) as well as words ending in consonants other than alveolar and velar nasals (*cat*, *desk*). Evidence from Hawkins and Liszka (2003) showed perfect production by their Chinese subjects of past participle endings (100%) and a high rate of word-final [t/d] in monomorphemic words (82%). Similarly, Hansen's (2001) study of the speech samples of three Chinese informants (TOEFL590-617) found a variety of words with single, two-member and three-member coda in two data sets over a five-month period, many of which contained [t/d].⁴ Obviously, L1 Chinese syllable structure cannot be the source of variable marking of English simple past tense.⁵

From a parametric perspective, Hawkins (2000) and Hawkins and Liszka (2003) observed that in Chinese, the parametric feature [\pm past] is not selected by Chinese language. Indeed, unlike English, German and Japanese, Mandarin Chinese has no grammaticalized tense. It does not use verb affixes to signal the relationship between 'the time of the occurrence of the situation and the time that situation is brought up in speech' (Li & Thompson 1981: 184). The concept of 'past' is not marked morphologically, but expressed through lexical means and pragmatic contexts. Hawkins and Liszka (2003:36) claimed that this L1 feature is subject to the maturational constraint and 'will not be accessible in later L2 acquisition.' Evidence for this claim came from their study in which they compared the L2 English past tense marking by advanced end-state L2 English learners of Chinese (2), Japanese (5) and German (5) backgrounds who were Masters students at a university in UK. The results from an oral production task showed that the Chinese informants supplied the past tense marker at a much lower rate (62.5%) than their Japanese (91.9%) and German (96.3%) informants although all three groups were similar in a written test.⁶ While these results were consistent with Beck's (1997) conclusion that there was no deficit in the L2 morphological

4. Patty, on the other hand, showed a deletion rate of over 97% for monomorphemic words ending in [t/d], which is consistent with her non-production of past tense *-ed* in the past context (Lardiere 2003:180). Patty had a complex linguistic background. Considering that one of her L1s is Cantonese, a dialect of Chinese that allows the coda to be nasal stops as well as corresponding but unreleased bilabial, alveolar and velar stops [b, d, g] (Deng 1992), it is possible that Patty produced the past *-ed* at a rate higher than 5.80% but the production was not 'heard' because it was phonetically unreleased.

5. Incidentally, the Japanese syllable structure does not permit a consonantal coda either. Yet the simple past tense marking by Japanese-English learners was not compromised according to Hawkins and Liszka (2003).

6. The informants were international postgraduate students in the Masters programs of various academic disciplines at the University of Essex. The two Chinese informants' undergraduate major at their Chinese universities was unknown (Hawkins p.c. 2010).

component, the comparatively 'low' level of past tense marking by the Chinese informants, according to Hawkins and Liszka (2003), was due to the L1 parametric features of [past], which cannot be reset by adult L2 learners.

Hawkins and Liszka's conclusion was challenged by Lardiere (2003, 2008), who argued that even if a particular parameter (or feature) such as [+past] or [+plural] existed in two languages, the feature may vary greatly in complexity and learners must figure out 'the obligatory or optional conditions and restrictions on [the] overt expression of the feature' (Lardiere 2008:5). The English past tense expresses a number of obligatory distinctions: (1) past vs nonpast; (2) irrealis mood vs non-irrealis mood; (3) verb vs non-verb; (4) regular vs irregular verbs; and (5) past vs the pragmatic-driven 'historical present' (Jacobs 1995; Pinker 2007). In other words, the past tense marking in English encodes not only formal morphosyntactic features, but also semantic and pragmatic functions – a typical case of one-to-many mapping between form and function. Therefore, it was not clear, according to Lardiere (2003), how the parameter argument could work since 'there isn't a single overt morphological reflex that encodes or divides up exactly the same bunch of stuff...in exactly the same way' (p.187).

Lardiere (2008) proposed that the variable past tense marking by Chinese-English learners was due to their imperfect 'morphological competence,' or 'the knowledge of precisely which forms go with which features' (Lardiere 2008:4). However, evidence from the emails of her informant (Patty) showed that her imperfect L2 morphological competence and knowledge seemed limited to the oral production only, since Patty inserted *-ed* in her emails at a rate of 76.92% (Lardiere 2003), in contrast to 5.80% in her speech. Similarly, Beck (1997) and Hawkins and Liszka (2003) found no significant differences in written tasks between the Chinese group and the control group. Thus the L2 'morphological competence' as conceptualized in Lardiere (2008) in terms of 'declarative knowledge' cannot fully account for the variable past tense marking in the oral production.

So the question remains: why do advanced L2 learners of Chinese background have trouble inflecting simple past tense at a near-native rate? In this study, we wish to continue the debate through two hypotheses:

Hypothesis 1: Rigorous training is key to native-like simple past tense marking by Chinese-English speakers. We define 'rigorous training' in terms of the English language programmes offered for English-major (EM) students at the prestigious universities in China. 'Prestigious universities' are those so-called '211' and '985' universities designated by the Chinese Ministry of Education.⁷ English major

7. Of over 2,000 universities in China, 122 are '211' universities and 44 are '985' universities. Traditionally prestigious and well-known, these universities enjoy priority funding, quality staff, academic rigor, as well as development opportunities for both staff and students.

students at these universities receive rigorous and professional English language training not available to non-English major students (NEM) or students at non-prestigious universities. We hypothesize as a result of such quality training, the EM learners will inflect verbs for the past tense at a similar rate as the Japanese and German learners in Hawkins and Liszka's (2003) study.

While formal instruction has been shown to have a definite advantage in the L2 learning outcome, research typically compares tutored and non-tutored learners, types of instructions, and the length of instruction (for an overview, see Doughty 2003). Few studies examined the relationship between the learning experience and the learning outcome of end-state learners. We hypothesize that learners who have gone through top-notch language programmes in which skill training is emphasized and properly delivered do not display variable marking in the past tense.

Hypothesis 2: Rigorous training discourages 'bad choices' being made by the learner. The 'bad choice' hypothesis was proposed in Pienemann (1998) to account for the IL variational features. According to Pienemann (1998), learners use a variety of ways to deal with production and developmental problems. Omission is one of them. Omission of copula and inflectional morphemes are well documented in IL studies. While this enables the learner to meet their immediate communicative needs and even allows them to progress along the developmental path, it has a flow-on effect in the subsequent IL development. Pienemann (1998:326) pointed out that 'a learner with the most highly simplifying features also displays all other variable features.' In other words, simplified features of IL are connected and remain constant along the IL developmental course.

To test our H_2 that rigorous language training suppresses 'bad choices,' we analyzed the plural marking -s to see if indeed, there was a match between the simple past tense marking and the plural -s marking.⁸ If our H_1 were true (rigorous training is key to native-like simple past tense marking), our H_2 should also be true (rigorous training discourages 'bad choices' being made). Variable markings, or 'bad choices,' do not materialize in the IL of learners from top-notch English language programmes.

2. The study

2.1 Informants

We invited as informants 9 highly advanced end-state Chinese-English speaking professionals who had studied and graduated from prestigious education institutions

8. We thank one of the reviewers for this suggestion.

in China.⁹ These 9 academic staff (5 male and 4 female) were lecturers, senior lecturers and associate professors at an Australian university, teaching mathematics (1), physics (1), IT (1), human resources (1), law (2), education (2), and academic skills (1). They had completed their Bachelor's degrees in (mainland) China in the 1980s and 1990s. Five of them majored in English (EM), and four in non-language disciplines (NEM). Eight of them obtained their Doctorate degree in Australia, USA, and Canada, and one was in the process of completing her Ph.D. dissertation in Australia. Aged in their late 30s and early 60s, they had lived and worked outside China (US, Australia, Hong Kong, Europe, Singapore) for a minimum of 10 years. Their English proficiency, formally assessed through IELTS or TOEFL prior to their study outside China, was above TOEFL500 for the four non-English major informants, and above TOEFL600 or IELTS7.0-7.5 for the five English major informants. Additional evidence of their being advanced L2 English speakers was their current occupation – university academic staff – which requires a high level of language proficiency in addition to professional knowledge in their discipline areas.

Eight of the 9 informants studied at prestigious universities in China. Although one EM informant's university was not in that category, the English language course in her high school was of a similar standard.¹⁰

2.2 Data collection

The L2 English speech samples were elicited through interviews in English. Each interview lasted over 50 minutes. The informants were asked to recall their English language learning experience back in China, including their English classes, extra-curricular activities, motivations and feelings about the training at the university in general. They were also asked about their experience outside China, such as their Ph.D. studies, and life as students and lecturers. These conversation topics were all concerned with past events, and contained a multitude of obligatory contexts for the use of past tense verb forms including *-ed*. We also interviewed a native speaker of English as a control. The native speaker indeed treated these topics as past events, recounting her own language learning experience many years ago in the past tense.

The interview style was conversational, similar to a story-telling event. We guided the conversation and interacted with the informants but refrained from interrupting their story-telling sequence so long as the topics were concerned with past events.

9. Written informed consent was obtained from the informants.

10. This informant stated that she had acquired nearly all her English knowledge and skill in high school.

3. Data analysis and results

Hypothesis 1: Rigorous training is key to native-like simple past tense marking.

The audio recording of all 9 interviews were transcribed and cross-checked by the researchers for accuracy. Particular attention was paid to the regular verbs which required the *-ed* ending. One of the difficulties in identifying the obligatory context for the past tense marking was determining the informant's intention (what they meant to say). When the discourse context in which the informants' utterances failed to provide sufficient cues to determine the temporal reference of the utterance, the utterance was excluded from analysis. Following Lardiere's (1998a) exclusion criteria, we also did not include the following:

1. A past situation context where the situation still holds true in the present and therefore a present tense temporal reference is equally possible (e.g. *She's maybe ten years old*)
2. Formulaic expressions
3. Instances where the past and non-past forms are similar (e.g. *put*)
4. Quotations or reported speech
5. Contexts in which the past tense inflection is adjacent to homophonic stops (e.g. *We exchanged ed diary. I stoped talking.*)
6. Utterances followed immediately by spontaneous self-correction

We calculated the suppliance rate of the past tense *-ed* in the obligatory contexts of four verb categories:

1. All verbs
2. Thematic verbs or lexical main verbs: *drive, talk, eat, study*
3. Regular verbs: *talk, study*
4. Irregular verbs: *drive-drove, eat-ate*

Table 1 shows individual informants' past tense marking in obligatory contexts. Except for the regular verb category in EM04, the overall data density is high. Figure 1 shows the overall results. The past *-ed* suppliance, although the lowest among the four verb categories, nevertheless reached 61%, a comparable rate to that of Hawkins and Liszka's (2003) Chinese informants (62.5%). Irregular verbs, on the other hand, had the highest marking rate (73%). Both were much higher than Patty's (5.8% on regular verbs, 40% on irregular verbs). When the regular past *-ed* suppliance of the EM and NEM informants was analysed separately (Figure 2), we saw that the EM group outperformed the NEM group by a large margin (71% and 47% respectively). Figure 3 shows that our EM informants also

Table 1. Past tense marking

Informants	Verb type	Number of obligatory context	Past tense suppliance	Suppliance rate(%)
EM01	Irregular	73	66	90.41
	Regular	46	34	73.91
	Thematic	119	100	84.03
	All Verb	228	181	79.39
EM02	Irregular	61	55	90.12
	Regular	49	36	73.47
	Thematic	110	91	82.72
	All Verb	179	142	79.33
EM03	Irregular	32	29	90.63
	Regular	22	13	59.09
	Thematic	54	42	77.78
	Verb	129	85	65.89
EM04	Irregular	37	29	78.38
	Regular	5	4	80
	Thematic	42	33	78.57
	All Verb	135	121	89.63
EM05	Irregular	63	41	65.08
	Regular	34	24	70.95
	Thematic	97	65	67.01
	All Verb	202	129	63.86
NEM06	Irregular	74	68	91.89
	Regular	50	36	72
	Thematic	124	104	83.87
	All Verb	231	183	79.22
NEM07	Irregular	50	34	68.00
	Regular	14	10	71.42
	Thematic	64	44	68.75
	All Verb	155	110	70.97
NEM08	Irregular	110	49	44.54
	Regular	51	10	19.60
	Thematic	161	59	36.64
	All Verb	314	92	29.20
NEM09	Irregular	107	45	42.06
	Regular	57	15	26.32
	Thematic	164	60	36.59
	All Verb	364	111	30.49

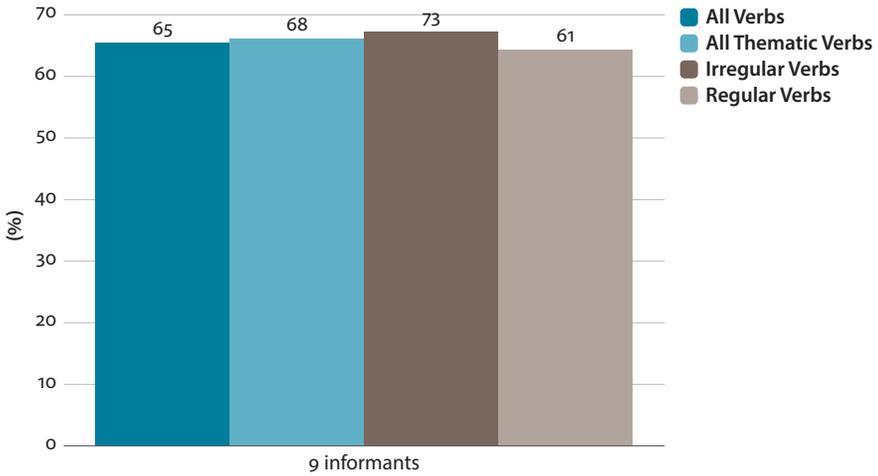


Figure 1. Overall past tense marking (%)

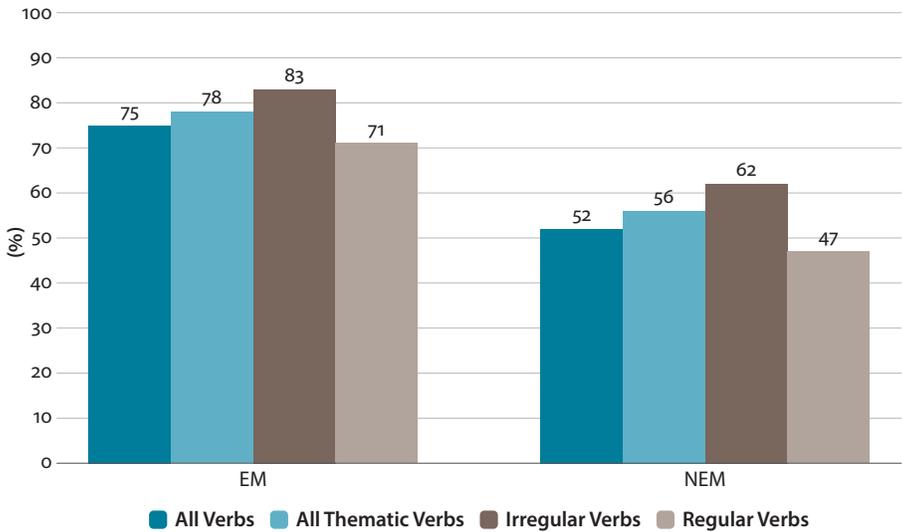


Figure 2. Past tense marking by EM and SM (%)

outperformed Hawkins and Liszka's Chinese informants (71% vs. 63%), but not their German and Japanese informants (96% and 92%).

The NEM group displays large individual variations. Figure 4 shows that two of the NEM informants (06, 07) performed at the level of the EM informants, with above 70% past *-ed* suppliance rates (72%, 71%). The other two NEM

informants (08, 09) had low past marking rates (20%, 26%) although still much higher than Patty.

In sum, as a group, the results corroborate Hawkins and Liszka's (2003) findings, and the suppliance rates of *-ed* in both studies are higher than Patty (Lardiere 1998a/b). Our EM informants marked the regular past tense more consistently and at a higher rate than both our NEM informants and Hawkins and Liszka's Chinese informants although they did not reach the level of the German and Japanese informants.

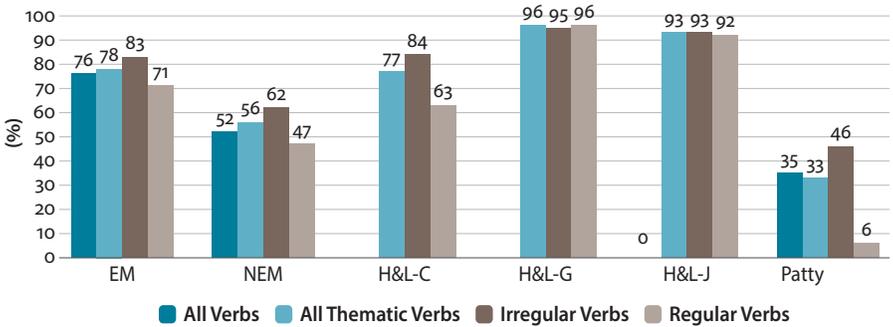


Figure 3. Past tense marking by EM, SM, Chinese (H&L-C), German (H&L-G), Japanese (H&L-J), and Patty (%)

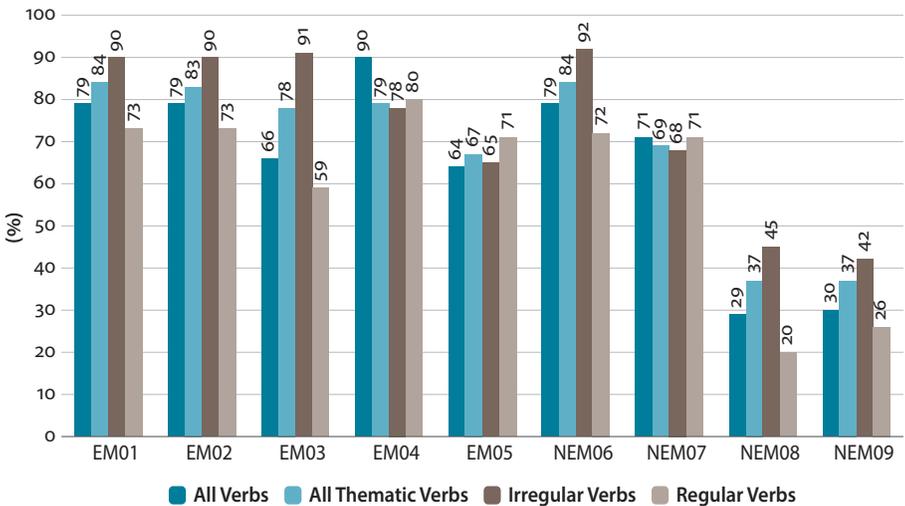


Figure 4. Individual results of past tense marking (%)

Hypothesis 2: Rigorous training discourages 'bad choices' being made.

To test H_2 , we analyzed the plural marking in the data to see if there was any connection between the past marking and the plural marking. This is because according to Processability Theory, the 'bad choices' made in the IL are not isolated

instances. We identified two types of plural contexts: the lexical plural (*I like oranges*) and the phrasal plural (*these/two oranges*) which requires agreement. According to the processing hierarchy for L2 English (Pienemann 2005: 24), the lexical plural is situated at the same developmental stage as the simple past *-ed* while the phrasal plural is one stage higher. If H_2 were true, we should see a correlative trend between the past tense marking and the plural marking at both group and individual levels. Furthermore, the EM group should display a homogeneous characteristic.

The obligatory plural contexts and the plural marking in these contexts were noted in the transcripts. Four of the 9 transcripts were double-checked by a native speaker of English for analytical accuracy. Table 2 shows the obligatory plural contexts, the suppliance of *-s*, and the suppliance rates. Figure 5 displays the marking rates by the EM and the NEM groups. Clearly, the EM supplied both plural markers at a higher rate than the NEM, and both groups performed better than their past tense marking, as shown in Figure 6. Similar to the past tense marking, NEM06 and NEM07 reached the level of the EM group, while NEM08 once more brought up the rear. As a group, the EM had a high level of homogeneity, as evidenced in a smaller range (87%–98% for the lexical *-s* and 87–96% for the phrasal *-s*).

Table 2. Plural marking

Informants	Plural types	Number of obligatory contexts	Number of plural suppliances	Suppliance rate (%)
EM01	Lexical	46	42	91
	Phrasal	47	41	87
EM02	Lexical	95	93	98
	Phrasal	61	58	95
EM03	Lexical	69	60	87
	Phrasal	42	38	90
EM04	Lexical	60	54	90
	Phrasal	25	24	96
EM05	Lexical	39	34	87
	Phrasal	32	28	88
NEM06	Lexical	98	94	96
	Phrasal	57	55	96
NEM07	Lexical	35	30	86
	Phrasal	22	21	95
NEM08	Lexical	42	26	62
	Phrasal	42	25	60
NEM09	Lexical	64	55	86
	Phrasal	49	43	88

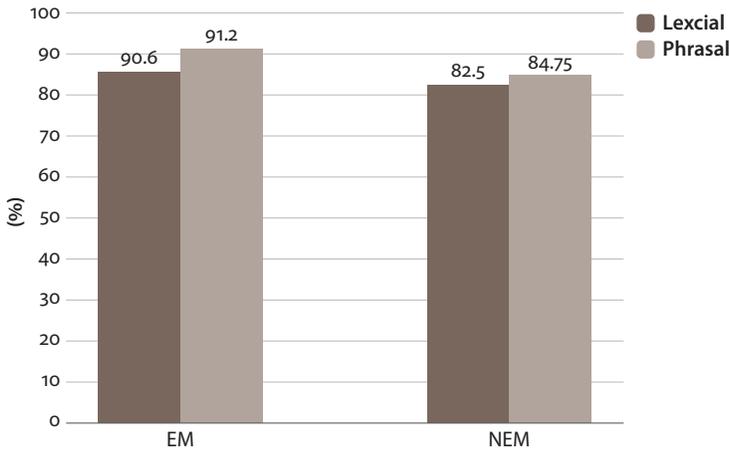


Figure 5. Plural marking by EM and NEM (%)

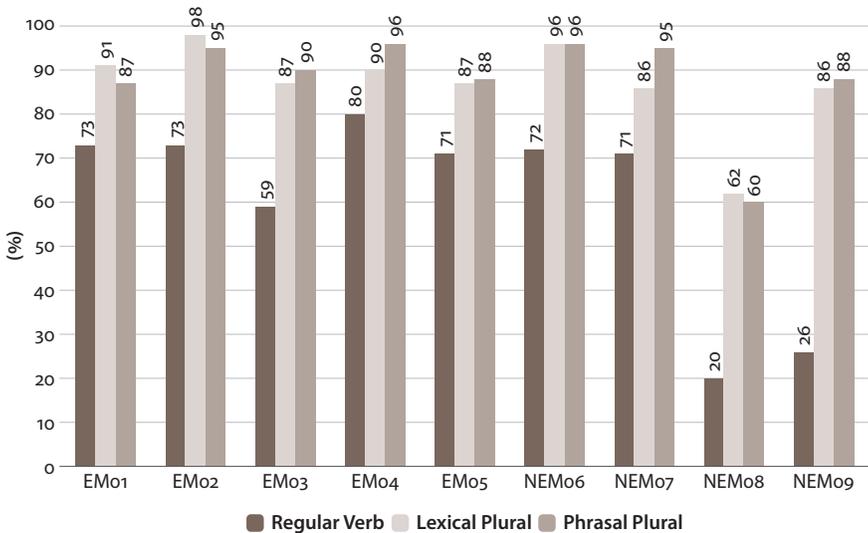


Figure 6. Individual results of past (-ed) and plural marking (%)

Indeed, the high level of the past tense marking was matched by a comparable high level of the plural marking in 7 informants. One NEM informant (08) was low in both. This suggests a connection between the features of IL, supporting the generative entrenchment claim in L2 learning as well as H_2 . The one exception seemed to be NEM09, whose low past marking (26%) was not duly reflected in his high plural marking (lexical 86%, phrasal 88%). In the following, we will discuss our findings in connection to the hypotheses, and the case of the two NEM high achievers.

4. Discussion

The consistent performance at a high level both as a group and at the individual level by the EM group supports H_1 : Rigorous training is key to native-like simple past tense marking by Chinese-English speakers. It is testimony to the value of a well-organized, all-round and rigorous training program. In the 1980s and early 1990s when our informants were university students in China, the English language teaching in China was characterized by the 'focus-on-formS' approach (Long & Robinson 1998),¹¹ with the emphasis on L2 grammatical and lexical knowledge. L2 accuracy overrode L2 fluency and communicative skills. In the universities, the quality of language programs, language teachers and classroom instruction varied greatly between EM and NEM programmes, and this was reflected in the quality of the curriculum and the competence of the teachers in terms of their L2 knowledge and skill to organize and deliver instructions. The mission of the English department of the prestigious universities was to produce language professionals for foreign affairs, translation and interpretation, international business, journalism, and tertiary institutions. The target proficiency level for the EM students at the end of their four-year study was native-like L2 linguistic and communicative competence. To this end, the EM curriculum contained a variety of courses with clearly articulated goals, and was delivered systematically to students throughout their degree programme. In addition to core language courses (listening, speaking, reading, writing and translation), EM students also took courses in English and American literature, western culture and society, international politics, and world history. These were usually taught in English, often by native English-speaking 'foreign experts' (*waijiao* 外教). The Chinese teaching staff in the English department were themselves highly proficient and often (near) native-like in English. Many of them had received education in missionary schools or spent time in the English speaking countries.

Classes for the EM were small, with 15 to 20 students per class. Classroom teaching was characterized by both focus-on-form and focus-on-formS (Long 1991; Long & Robinson 1998). Accuracy and fluency were emphasized and demanded equally. L2 input and practice took place both inside and outside the classroom, and students had access to English language resources such as native speakers, English language films, books, magazines, and international radio broadcasts. They also had more opportunities for the extensive application of their L2 knowledge and skill during their studies.

11. According to Long and Robinson (1998, also Doughty & Williams 1998), *Focus on formS* refers to the kind of instruction that focuses on the formal elements of language.

In contrast, NEM students were required to take 'General English' courses (*gonggong waiyu* 公共外语) only in the first two years of their four-year university study. Although compulsory, the language courses were peripheral to their discipline courses. The classroom instruction focused on L2 knowledge exclusively with an explicit emphasis on L2 grammar and vocabulary. General English classes were large. It was not uncommon to have a class of 50 to 200 students from various disciplines under one roof. The teaching staff was not required to have a high L2 proficiency as they were not required to teach in L2 English. Students rarely had the chance to see native speaker teachers, let alone being taught by them. Overall, general English courses in the academic life of the NEM students were not accorded the same status as those for the EM students. Table 3 is a summary of the key features of the language programs for EM and NEM students.

It seems clear from the sketch above, reported by the informants, that the differential training regime during the formative years of our informants' academic study was reflected in the end-state of their L2 English, in particular, in the past tense marking.

Table 3. Programmes for English major and non-English major in Chinese universities (prior to 1999)

English Programme	English major (EM)	Non-English major (NEM)
Length (years)	4	2
Target proficiency level	Native-like	Not explicitly specified
Focus	Comprehensive L2 knowledge and functional skills	Grammar and vocabulary
Class size	12–20 students	30–200+ students
Instruction	15–25 hr/week	4–5 hr/week
Language of instruction	L2 (English)	L1 (Chinese)
Instruction format	Lecture (teacher-front), tutorial, pair/group work	Lecture (teacher-front)
Staff	Chinese, English native speakers	Chinese
Curriculum	Variety of courses in and about L2	General English

How do we account for the consistent high level of tense and plural marking by two of the NEM informants (e.g., NEM06, NEM07)? According to the stories of our NEM informants, what had not been provided by the English language programme and classroom instruction was compensated for by a rigorous regime of self-training driven by an extraordinarily high level of motivation. All four NEM informants reported similar classroom experience in their undergraduate studies: teacher-centered pedagogy, exclusive focus on grammar and vocabulary,

grammar-translation teaching methods, large classes, and instruction in L1. They also reported similar experiences outside the classroom: actively seeking out opportunities to receive input by and interact with native-speakers of English, participating in extra-curriculum activities to practice and use English, and implementing an intensive and continuous self-training regime. Strongly goal-oriented and highly motivated, they made a huge effort working towards the level of English proficiency required to study abroad. All four of them also reported a drastic change in their postgraduate studies in China: more focus on speaking and listening skills, more time spent on self-study, and seeking every opportunity to use English, for example, attending seminars given by international scholars, watching English-speaking films, practicing English in the 'English corner'¹² on weekends, listening to VOA or BBC. Hard work paid off, but the extent and quality of the outcome varied, indicating that without a quality programme that formally organizes and delivers rigorous and consistent training professionally, the learning outcome varies greatly. The possibility of 'bad choices,' i.e., forming hypotheses that allow the acquisition of a simplified form to meet immediate communication needs, increases.

Indeed, if we examine the suppliance of the plural marking in the data, we see that those informants who did not opt for the omission option for the simple past tense marking also did not do so for the plural marking (see Figure 4 and Figure 6: the EM group, NEM06 and NEM07). The 'bad choice' was not entertained by these informants. NEM08 was low in both, indicating a 'bad choice' scenario. The only exception was NEM09, whose past tense marking was low but whose plural marking was not. Overall, the developmental profiles of the informants in the two inflectional morphemes under study seem in line with the 'bad choice hypothesis' or 'developmental dynamics' discussed in Pienemann (1998: 326–327), 'learners who do not progress far along the developmental axis after a long period of exposure have developed a highly simplifying variety of the L2.' Previous studies such as Clahsen, Meisel and Pienemann (1983, cited in Pienemann 1998) on L2 German and Lardiere (1998a/b) on L2 English support the view. In the L2 German study, it was found that despite more than 7 years of exposure, a group of learners exhibited highly simplified features in their L2 German below Stage X+2 (verb separation). Similarly, Patty in Lardiere (1998a/b) supplied a mere 4.5% 3rd person -s and 34.5% past tense marking despite her 10+ years of living and working in the US. On the other hand, similar to our informants who did not make 'bad choices,' SD, an adult Turkish-speaking learner of

12. English corner: a spontaneous gathering to practice English in parks. Anyone and everyone can participate. It started in the late 1970s and still exists in some cities today.

L2 English, consistently supplied a high level of 3rd per -s (Time1: 78%, Time2: 81.5%) and the past tense¹³ (Time1: 85%, Time2: 76%) after living in Canada for 10 years (White, 2003).

The findings from previous research as well as our study indicate that the variable marking by L2 English learners of Chinese background is mainly confined to the simple past tense *-ed* in the oral form. From the processing perspective, the past *-ed* and the lexical plural *-s* are both lexical morphemes, and therefore require the same processing procedures (Pienemann 1998). Indeed, they had been successfully acquired by all the informants, as measured by the 'emergence criterion' (Pienemann 1998). However, in terms of the ultimate attainment as measured by the accuracy criterion, the Chinese informants in all the studies fell short. Since online processing skills (Pienemann 1998) and L2 morphological competence (Lardiere 2008) apply to all L2 learners regardless of L1 parametric settings, German and Japanese learners face the same processing issue as Chinese learners when learning L2 English. Yet, the German and Japanese learners in Hawkins and Liszka (2003) were able to supply the past tense *-ed* more consistently at near-native rates, and Chinese informants were not. It seems the one factor that sets the Chinese apart from German and Japanese is the presence of the past tense in German and Japanese. This appears to give an advantage to adult L2 English learners of German and Japanese backgrounds. It seems that the past tense morphology, if exercised in the L1 through first language acquisition, remains and assists with L2 past tense learning. This may offer an additional explanation for the near-native performance of the German and Japanese but not the Chinese informants.

5. Conclusion

In this study, we investigated the variable past tense marking in the L2 English of Chinese learners. Although (formal) tutoring has long been proved to be superior in second language acquisition than non-tutoring, the level of ultimate attainment that a well-organized and professionally executed language programme can achieve has not been documented for this group of learners. Our own study supports the argument that a rigorous training regime indeed enables a uniformly high level of skill development and discourages 'bad choices' being made.

By way of conclusion, we would like to offer some suggestions for foreign language teaching. First, Chinese language teachers (and learners) should be made

13. It seems there was no separation between regular and irregular verbs.

aware of the past marking issue.¹⁴ A lack of awareness, compounded with a mediocre training programme, tends to result in a high level of variable past marking in the end-state. Pedagogical intervention should take place early, covering a range of (past) tense situations and focusing on regular verbs through 'Processing Instruction' (VanPatten 2007), focus-on-form and focus-on-forms approaches (Long & Robinson, 1998).

Given the learning experience of our informants, we would like to propose a shorter but intensive training programme for NEM students similar to that enjoyed by our EM informants, because it may achieve what a two-year General English course is unable to do. Without it, the NEM students must invest time and effort to train themselves in order to attain a high level of L2 skill, and this may not be achievable for everyone.

6. Acknowledgement

The early versions of this paper were presented at the 10th International Symposium on Processability Approaches to Language Acquisition (University of Western Sydney 19–21 September 2010), the 1st Conference on the National Tertiary English Language Pedagogy Conference (Shanghai Jiaotong University, 26–28 November 2010), and the Language and Linguistics Cluster Seminar (University of Canberra, May 2011). We thank the audience for their comments. We would also like to express our gratitude to our informants, whose graciousness and generosity to share with us their learning experience are greatly appreciated. We thank Eileen Bower for her assistance with data collection and some data transcription. Finally, we have benefitted from the guidance of the reviewers and the editors and we thank them all. The interpretation and errors that remain are ours.

References

- Beck, M. (1997). Regular verbs, past tense and frequency: Tracking down a potential source of NS/NNS competence differences. *Second Language Research*, 13(2), 93–115.
doi:10.1191/026765897670780840
- Clahsen, H., Meisel, J., & Pienemann, M. (1983). *Deutsch als Zweitsprache. Der Spracherwerb Ausländischer Arbeiter*. Tübingen: Narr.
- Deng, S. (1992). 《实用粤语教程》。上海：海文音像出版社。[Practical Cantonese Textbook]. Shanghai: Shanghai Audio-Visual Press.

14. Many (Chinese) teachers and students, including ourselves, were unaware of this before this study.

- Doughty, C. (2003). Instructed SLA: Constraints, compensation, and enhancement. In C. Doughty & M. Long (Eds.), *The handbook of second language acquisition* (pp. 256–310). Oxford: Blackwell. doi:10.1002/9780470756492.ch10
- Doughty, C., & Williams, J. (1998). *Issues and terminology*. In C. Doughty & J. Williams (Eds.), *Focus on Form in Second Language Acquisition* (pp. 197–261). Cambridge: Cambridge University Press.
- Eubank, L., & Grace, S. (1998). V-to-I and inflection and non-native grammars. In M. Beck (Ed.), *Morphology and its interface in L2 knowledge* (pp. 69–88). Amsterdam: John Benjamins. doi:10.1075/lald.19.05eub
- Hansen, J. (2001). Linguistic constraints on the acquisition of English syllable codas by native speakers of Mandarin Chinese. *Applied Linguistics*, 22, 338–365. doi:10.1093/applin/22.3.338
- Hawkins, R. (2000). Persistent selective fossilization in second language acquisition and the optimal design of the language faculty. *Essex Research Reports in Linguistics*, 34, 75–90.
- Hawkins, R., & S. Liszka (2003). Locating the source of defective past tense marking in advanced L2 English speakers. In van Hout, R., A. Hulk, F. Kuiken, & R. Towell (Eds.), *The interface between syntax and lexicon in second language acquisition* (pp. 21–44). Amsterdam: John Benjamins. doi:10.1075/lald.30.03haw
- Jacobs, R. (1995). *English syntax: A grammar for English language professionals*. Oxford: Oxford University Press.
- Lardiere, D. (1998a). Case and tense in the ‘fossilized’ steady state. *Second Language Research*, 14(3), 1–26. doi:10.1191/026765898674105303
- Lardiere, D. (1998b). Dissociating syntax from morphology in a divergent L2 end-state grammar. *Second Language Research*, 14(3), 359–375. doi:10.1191/026765898674105303
- Lardiere, D. (2000). Mapping features to forms in second language acquisition. In J. Archibald (Ed.), *Second language acquisition and linguistic theory* (pp. 102–129). Malden, MA: Blackwell.
- Lardiere, D. (2003). Second language knowledge of [\pm past] vs. [\pm finite]. In *Proceedings of the 6th generative approaches to second language acquisition conference (GASLA)* (pp. 176–189). Somerville, MA: Cascadilla Proceedings Project.
- Lardiere, D. (2008). Feature-assembly in second language acquisition. In Licerias, J., H. Zobl, & H. Goodluck (Eds.), *The role of formal features in second language acquisition* (pp. 106–140). New York, NY: Lawrence Erlbaum Associates.
- Li, C., & Thompson, S. (1981). *Mandarin Chinese: A functional reference grammar*. Berkeley, CA: University of California Press.
- Li, Y.-H. A. (1990). *Order and constituency in Mandarin Chinese*. Dordrecht: Kluwer. doi:10.1007/978-94-009-1898-6
- Long, M. (1991). Focus on form: A design feature in language teaching. In K. de Bot, R. Ginsberty, & C. Kramsh (Eds.), *Foreign language research in cross cultural perspective* (pp. 39–52). Amsterdam: John Benjamins. doi:10.1075/sibil.2.07lon
- Long, M., & Robinson, P. (1998). Focus on form: Theory, research and practice. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 15–41). Cambridge: Cambridge University Press.
- Pallotti, G. (2007). An operational definition of the emergence criterion. *Applied Linguistics*, 28, 361–382. doi:10.1093/applin/amm018
- Pienemann, M. (1998). *Language processing and second-language development: Processability Theory*. Amsterdam: John Benjamins. doi:10.1075/sibil.15

-
- Pienemann, M. (2005). An introduction of Processability Theory. In M. Pienemann (Ed.), *Cross-linguistic aspects of Processability Theory* (pp. 1–60). Amsterdam: John Benjamins.
doi:10.1075/sibil.30.03pie
- Pinker, S. (2007). *The stuff of thought*. Victoria: The Penguin Group.
- Prévost, P., & White, L. (2000). Missing surface inflection or impairment in second language acquisition? *Second language Research*, 16, 103–133. doi:10.1191/026765800677556046
- Vainikka, A., & Young-Scholten, M. (1994). Direct access to X'-Theory: Evidence from Korean and Turkish adults learning German. In T. Hoekstra & B. D. Schwartz (Eds.), *Language acquisition studies in generative grammar* (pp. 265–316). Amsterdam: John Benjamins.
doi:10.1075/lald.8.13vai
- VanPatten, B. (2007). Input processing in adult second language acquisition. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition: An introduction* (pp. 115–136). Mahwah, NJ: Lawrence Erlbaum Associates.
- White, L. (2003). Fossilization in steady state L2 grammars: Persistent problems with inflectional morphology. *Bilingualism: Language and Cognition*, 6, 129–141.
doi:10.1017/S1366728903001081

Acquisition as a gradual process

Second language development in the EFL classroom

Jana Roos

Paderborn University

This chapter explores the potential of communicative tasks with a developmentally moderated focus on form to promote the acquisition of grammatical features in the EFL classroom. Task-based language teaching in combination with a focus on form is discussed as a methodological approach that can provide learners who are developmentally ready to acquire a structure with opportunities to use it spontaneously and productively in different contexts. Examples of task-based interactions between German learners of English at different levels of acquisition will be presented that illustrate how such tasks can be used to stimulate the acquisition process in the classroom.

1. Introduction

Substantive research in the field of Second Language Acquisition (SLA) shows that learners of a foreign language follow predictable stages in their interlanguage development, in both natural and classroom contexts (e.g. Meisel et al. 1981; Pienemann 1998; Lenzing 2013). This clearly has important implications for the teaching of languages. On the one hand, an invariant developmental sequence implies that there are limits on the effect of instruction on acquisition. However, there is also evidence that instruction addressing a learner's current stage of interlanguage development may have beneficial effects on the acquisition process (Pienemann 1989; VanPatten & Williams 2007; Pienemann & Keßler 2011). In any case, a conclusion that appears obvious is that these developmental sequences need to be taken into account if foreign language instruction is to be effective and tuned to the learners' needs.

For the language teacher, this inevitably leads to three fundamental questions: Which grammatical structures (*what*) should be taught at which point in time (*when*), and *how* can this be realised from a pedagogical perspective? (Doughty & Williams 1998a:6) These three guiding questions serve as an outline for this chapter. The first two questions, what to teach and when to teach it, have

been widely discussed and have set the ground for numerous studies in SLA (e.g. Lightbown 1998; Roos 2007). They will be looked at in the first part of this chapter, which briefly summarizes implications of the Teachability Hypothesis (Pienemann 1989) and the concept of developmental readiness for language teaching. The two questions are closely connected because the learnability and thus the teachability of linguistic structures require that a learner is developmentally ready to process these structures, i.e. has developed the respective processing mechanisms (Pienemann 1989; 1998).

The second part of the chapter will deal with question number three and thus with ways of promoting language learning in the classroom that take the learner-internal syllabus into account. Here, task-based language teaching in combination with a focus on form will be discussed as a methodological approach that provides opportunities for the use of specific linguistic features in a communicative classroom context (Ellis 2003; Long & Robinson 1998; Long 2011). The underlying idea is that tasks, which focus on aspects of form that are learnable, can by design support the acquisition process and thus play a key role in a developmentally moderated approach to foreign language teaching. In the following, I will refer to these tasks as “tasks with a developmentally moderated focus on form”. Examples of how learners interact while working with such tasks will be presented, in order to illustrate their potential to contribute to second language development in English as a foreign language (EFL) classrooms.

2. Language teaching and developmental readiness

With regard to the timing of instruction, and the question what to teach and when, traditional approaches to foreign language teaching are based on the idea that language learning is a linear process. The basic principle is that structures that are perceived to be simple are taught before complex or difficult ones. This principle is accompanied by an assumption that items are learned in the order in which they are taught. These ideas are reflected in a synthetic syllabus, in which the target language is segmented into discrete linguistic items, suggesting that language learning is a process of accumulating these items one after another (Lightbown & Spada 2013; Long 2011).

Research has shown, however, that learners do not learn isolated grammatical structures one after another. On the contrary, second language acquisition is “a gradual and dynamic process” (Ellis 2009:237), in which “learners rarely, if ever, exhibit sudden categorical acquisition of new forms or rules [...]” (Long & Robinson 1998:16). Irrespective of which structures are introduced at which point in time, this research tradition has demonstrated that learners follow a

natural, irreversible order of developmental stages. In Processability Theory (PT) (Pienemann 1998), this staged development in second language acquisition is explained in terms of constraints on language processing and “the sequence in which processing procedures become available to the language learner” (Liebner & Pienemann 2011: 65).

If we ask the question what should be taught when, again from this psycholinguistic perspective, as Pienemann (1984; 1989) initially did in the Teachability Hypothesis, this leads to “the unavoidable conclusion that [...] [s]tudents do not – in fact, cannot – learn (as opposed to learn about) target forms and structures on demand, [...] but only when they are developmentally ready to do so” (Long 2011: 378). This means that a learner at stage 1 does not have the necessary prerequisites to acquire structures of stage 3, but may well benefit from instruction focusing on structures from the next developmental stage (stage 2). Thus, studies carried out within the framework of the Teachability Hypothesis show that stages of acquisition cannot be skipped through instruction, but that targeting learnable features can facilitate the acquisition process (Ellis 1989; Pienemann 1989; Mansouri & Duffy 2005; Keßler 2006). This finding not only confirms that it “pays” to take learners’ developmental readiness into account in the teaching process, it also adds a new and beneficial dimension with regard to the timing of instruction.

Having looked at the questions of *what* and *when* in the language teaching process, what has been illustrated about the way languages are learned can serve as a basis for examining the question of *how* with regard to appropriate types of pedagogical interventions. In the next part of the chapter, task-based language teaching will be looked at as a promising approach since it has been shown to be compatible with developmentally moderated approaches to language teaching (Keßler & Plessner 2011; Keßler, Liebner, & Mansouri 2011).

3. Task-based interaction in the classroom

The possibilities a task-based approach offers for the teaching of foreign languages, and its potential to support processes of second language acquisition, have been discussed from many different theoretical and practical perspectives (e.g. Ellis 2003, 2009; Mackey 1999). While there are a large number of different definitions of “task” in this context, one that encompasses most characteristics commonly attributed to tasks is offered by Ellis (2009: 223). Ellis basically describes a communicative task as a meaning-focused activity, which involves a need to convey information and enables learners to use the linguistic means available to them in order to achieve a clearly defined outcome. As outlined in Long’s Interaction Hypothesis (e.g. Long 1996), the information exchange, that takes place between learners, may

result in negotiation for meaning, elicit negative feedback and lead to interactional modifications that are needed in order to develop a shared understanding with their partner. This process of negotiated interaction is claimed to play an important role in language development, as Mackey (1999: 558) summarizes: “As linguistic units are rephrased, repeated, and reorganized to aid comprehension, learners may have opportunities to notice features of the target language.”

Thus, in foreign language teaching, tasks are generally viewed as means to promote the communicative and meaningful use of the target language. The opportunities for learner-based learning that task-based teaching provides and the goal-oriented interaction that tasks bring about as learners realise their communicative intentions can also be seen as a main reason why tasks have come to play a major role in the Common European Framework of Reference for Languages (CEFR) with its action-oriented approach (Council of Europe 2001: 157f.). Here, the learner’s role is described as that of a ‘social agent’ who completes tasks in different circumstances and makes use of, and at the same time develops, communicative competence, which comprises linguistic, pragmatic and sociolinguistic competence (Council of Europe 2001: 4; 9f.). Subsequently, tasks have begun to find their way into language teaching curricula and textbooks all over Europe (cf. e.g. Müller-Hartmann & Schocker-von-Ditfurth 2011).

With its primary focus on meaning, task-based language teaching contrasts with traditional form-oriented approaches. It is based on what is termed an analytic approach, which, by presenting target language samples, helps learners to “induce underlying rules and the meanings and functions of words” (Long 2011: 373) from linguistic input. Consequently, it attributes to the learner an active role in the acquisition process. While task-based teaching promotes the development of communicative competence as an overall goal, it can also support the acquisition of grammar, as Ellis (2009: 238) points out, because it “aims to create a context in which grammar can be acquired gradually and dynamically while at the same time fostering the ability to use this grammar in communication.” Even though there is an inherent focus on meaning and communication in a task-based approach, this does not rule out the possibility of integrating a focus on structures of the target language. This already indicates the relevance of task-based language teaching when it comes to considering the connections between the questions of what to teach, when and how.

4. Task-based language teaching and focus on form

Whereas there is a focus on discrete language forms in a synthetic approach and a focus on meaning in an analytic one, focus on form is “a methodological

principle in Task-Based Language Teaching” (Long 2000: 179), in which “learners’ attention is focused on form in the context of communicative activities” (Ellis 2009: 232–3).

“(D)uring an otherwise meaning-focused lesson, (...) learners’ attention is briefly shifted to linguistic code features, in context, when students experience problems as they work on communicative tasks, i.e., in a sequence determined by their own internal syllabuses, current processing capacity, and learnability constraints.”
(Long 2000: 179)

Long’s concept of focus on form is responsive in that it relates to situations in which a problem occurs or a spontaneous need arises. Whereas this focus-on-form approach is reactive and incidental in nature, others have expanded Long’s definition to include the possibility of providing a focus on form in predetermined ways. For example, Doughty and Williams (1998b) differentiate between a proactive and a reactive approach to focus on form, and Spada and Lightbown’s (2008) term ‘integrated form-focused instruction’ (FFI) also includes the possibility of determining a form focus in advance:

“In integrated FFI, the learners’ attention is drawn to language form during communicative or content-based instruction. (...) That is, although the form focus occurs within a communicative activity, the language features in focus may have been anticipated and planned for by the teacher or they may occur incidentally in the course of ongoing interaction.”
(Spada & Lightbown 2008: 186)

Research has shown that an approach in which there is an intentional focus on language form, can be effective and promote second language development, e.g. in classrooms in which the emphasis of the teaching is on content and meaning rather than on form, such as immersion classrooms (e.g. Doughty & Varela 1998; for an overview of studies see e.g. Doughty & Williams 1998; and Norris & Ortega 2000). A form-focused approach can also provide a context that allows learners to notice target features in the input and the interaction they are engaged in. This also relates to Schmidt’s (1990) ‘noticing hypothesis’ and the idea that getting learners to attend to forms in the input contributes to acquisition.

In her two reviews of research on the effects form-focused instruction (1997 and 2010), Spada also addresses the question of whether there is an optimal time to provide form-focused instruction. She comes to the conclusion that providing instruction at the ‘right time’ can be effective, but points to the fact that little research has been done to date that links the idea of developmental readiness with form-focused aspects of instruction and that “the psycholinguistic timing issue” has not been addressed at all after the late 1990s (Lightbown 2010: 229). Adding the dimension of developmental readiness to Long’s original definition of focus

on form, Di Biase (2002) proposed a developmentally moderated focus on form, suggesting that the learners' attention should be drawn to forms for which they are developmentally ready. In a study with two classes of primary school learners of Italian as L2, he showed that this kind of instruction can be effective and speed up the acquisition process. In both classes, a focus on 'learnable' structures was provided, and in one of them, focus on form was additionally used as a feedback strategy. Learners in both groups progressed from stage 1 to stage 2 or even stage 3 in the processability hierarchy, which Di Biase attributes to the form-focused instruction they had received. The form-focused feedback which was applied in one of the groups seemed to have an additional positive effect, as the learners in this group showed a more consistent development and a more accurate use of the targeted structures.

By creating tasks which focus on particular language structures, also called focused tasks by Ellis (2003, 2009), opportunities for the use of specific linguistic features in a communicative context can be provided. In a study by Samuda (2001), for instance, this has been shown to have a positive effect on the learners' increased production of target features. With regard to the notion of processing constraints on teachability (Pienemann 1998), an approach combining task-based language teaching and focus on form can be taken a step further by placing the focus of such tasks on language forms for which learners are developmentally ready.

5. Tasks with a developmentally moderated focus on form

If the advantages of a combined task-based and form-focused approach are merged with what is known about the teachability and the learnability of grammatical features, the questions of what to teach and when can be linked to the question of how. It is important to point out here that, as Spada's reviews confirm, there is a lack of research "in which the timing of form-focused instruction has been manipulated in instructional materials and its effects examined in relation to L2 learning." (2011:229) PT provides a theoretical framework for a psycholinguistically-motivated selection of tasks with a developmentally moderated focus on form. This approach should not be mistaken for a proposal to teach language features in isolation. On the contrary, it is significantly different from a synthetic approach, in which the selection and the sequence of forms to be taught are only based on a formal linguistic description of the target language, or, as Doughty and Williams (1998b:198) put it, "on intuition."

Tasks that target specific language features are often developed and used as instruments for data collection and research, because the way learners deal with

such tasks and the language they produce when solving them, generally reveal a lot about their current processing of target language structures. For example, Mackey (1994; 1999) investigated the efficiency of different interactive tasks and showed that they were successful in eliciting certain targeted morpho-syntactic structures from learners. By identifying in the elicited data structures that are part of the developmental schedule for English as a second language outlined in PT, a learner's current stage of interlanguage development can be determined. For the teacher, this can serve as an important reference point in order to decide what can be taught next (Lenzing 2013; Keßler & Pienemann 2011; Roos 2007).

Approaches using tasks with a developmentally moderated focus on form not only for diagnostic purposes but also in the EFL-classroom, in order to promote the acquisition of structures that are learnable in the sense of the Teachability Hypothesis, have only recently begun to be discussed and explored. For example, Keßler, Liebner and Mansouri (2011) sketch the possibilities of tasks that can be used to test if a learner has acquired a structure and at the same time to practice the same structure in the classroom. In this context, the authors also point out the advantages of using communicative tasks in heterogeneous classrooms, because learners can use their own linguistic resources to solve them and so reveal that they are at different stages of developmental readiness. In an information-gap task learners might use "different structures from different levels within the PT hierarchy" in order to ask questions, in ways that correspond to their varied stages of acquisition (Keßler, Liebner & Mansouri 2011: 155). The following section will provide practical examples of how tasks with a developmentally moderated focus on form can be used in the EFL classroom.

6. Using tasks with a developmentally moderated focus on form in the EFL classroom

When using tasks with a developmentally moderated focus on form, learners have the opportunity to negotiate meaning through interaction, while being provided with natural contexts for the productive use of the targeted features. In the following, examples of interactions between German learners of English at primary and secondary level will be presented. The learners are engaged in different tasks focusing on two morphological structures that are part of the developmental hierarchy outlined in PT, namely plural -s and third person singular -s. This paper does not present a full analysis of the learner data. Instead, selected learner utterances are discussed in order to illustrate how tasks with a developmentally

moderated focus on form can be used in the classroom to scaffold interaction and to stimulate the acquisition process.¹

6.1 Using tasks with a developmentally moderated focus on form: Plural -s

In a combined cross-sectional and longitudinal study, Lenzing (2013) and Roos (2007) have shown that in primary school contexts, where oral production of the target language is often imitative and limited to the use of formulaic utterances, learners only gradually start to use target language structures productively (for an overview see also Lenzing & Roos 2012). A linguistic feature the young learners seemed to struggle with, even though it is acquired early in the PT hierarchy, is plural -s. This can be seen in the following utterances, in which a learner uses plural forms such as *Two eyes* or *Two ears* alongside *One legs* (examples taken from Roos 2007; Lenzing 2013). In the last utterance, the plural form of the noun is used in a context that requires the singular. A distributional analysis of the learner data carried out by Lenzing (2013) shows that the plural forms used in the other two examples were not generated by rules but were used as unanalysed chunks and had been stored as unanalysed entries in the mental lexicon. It is assumed that this is the result of the English language learning environment, in which opportunities to use the language productively outside of fixed dialogues were rare. This formulaic and lexical nature of foreign language learning in primary schools has also been described in studies by Engel et al. (2009) or Di Biase (2002). Tasks with a focus on plural -s can be used to provide learners who are developmentally ready to acquire the structure with opportunities to use it spontaneously and productively in different contexts while enabling teachers or others to identify those learners who are not ready even though they can complete such tasks in an alternative manner. Working with such tasks can not only initiate and support language development. When learners interact, this process and its dynamics even become observable through the language they use.

The examples below illustrate the interaction of learners in grade three (aged 8–9 years) after 2.4 ((1) and (2)) and 1.7 ((3) and (4)) years of instruction. It resulted from work with a picture-differences task with a focus on plural-s. The pictures depict a garden and a house with a number of different rooms and other elements, and learners had to find out about differences in the number of the various elements in their pictures. The first example (1) shows that the task provides numerous contexts for the production of the targeted structure and also for lexical and morphological variation:

1. The transcripts presented in this chapter come from unpublished data collected by Jana Roos.

- (1) C1: In the garden the tree and in the garden a two trees and on the trees sind (= are) apples.
 C2: I have on the trees not apples.
 C1: I have on the the living room one lamp nee (= no) two lamps, a one watching TV and a desk.
 C2: I have got four living rooms and in the living rooms sind (= are) TVs, abouts one TV and in the bathroom ehm three toilets.
 C1: I have ehm in the bathroom one toilet.

Examples (2) and (3) are based on a task in which the learners had to interact in order to find matching pairs of various elements in different numbers depicted on picture cards. The examples show that the task challenges learners in their use of plural forms and illustrate the online processing that takes places during language production. In Example (2), the learner manages to correct himself and uses the correct form in each of the second attempts:

- (2) C05 I have got one bike, six bike ... ehm ... six bikes.
 I have got one ball and five ball ... balls.

A similar process can be observed in Example (3). Here, the learner adds a plural -s to the nouns only after a pause, once even after an acknowledging comment by the interviewer (I) observing the interaction between the learners:

- (3) C07 Three dog
 I Okay.
 C07 ... s
 Two rat ... s. One apple.

In order to maintain a focus on meaning during the task-based interaction, the tasks were described according to their intended (content) outcomes and the learners were not told which feature they were supposed to produce. Still, a learner's question about the way to deal with the task in Example (4) reveals that he was aware of the fact that singular and plural forms play a role in the task.

- (4) C06 Ist das so, dass ... ehm, dass ein dog und mehrere dogs zusammenpassen?
 Is it the case that one dog and several dogs match?

All in all, the examples show that on the one hand, learners have choices about how to approach the task. On the other hand, the task also leaves room for the learners to use the linguistic repertoire available to them at that point, regardless of whether it is target-like or not, as can be seen in Examples (2) and (3). This results in more simplified language use in Example (3), where the learner uses noun phrases to convey the information needed, whereas the learner in Example (2) is already producing complete sentences with a subject-verb-object (SVO) order. Tasks focusing on other linguistic features can be designed and used in similar ways.

6.2 Using tasks with a developmentally moderated focus on form:

Third person singular -s

A morphological feature that is acquired later in the acquisition process is third person singular -s, located on stage 5 of the PT hierarchy (Pienemann 1998). The following examples illustrate the interaction of learners in grade six (aged 11–12 years) after four years of instruction. The examples resulted from work with two different information-gap tasks with a focus on third person singular -s.

In the first task (Examples (5) and (6)), the learners each hold information about a child from England that the partner does not have (for example about pets, hobbies, likes or dislikes). The pairs talk about the child and exchange missing information in order to complete the child's profile. In the second task, a picture-differences task (Example (7); see appendix), the learners need to exchange information in order to find out what a boy called John does regularly on different days of the week and to complete his timetable. Examples (5), (6) and (7) show that both tasks are effective and elicit the targeted structure. However, the same examples also reveal that the learners have different ways of dealing with the challenges imposed on them by the task and its focus on plural -s.

As in Example (2) above, the learner in Example (5) uses self-correction in order to produce a correct form. In a first step, he realises that he is uncertain about the correct verb to use and asks the interviewer for the English translation of the verb 'to bake'. Then he integrates it into a sentence firstly in the bare form and then by adding the correct inflectional morpheme. This can be seen as an indicator that the target structure is used productively in this context.

- (5) C04 Lucy ... Was heißt backen?
 Lucy ... What does bake mean (in English)?
 I Bake.
 C04 Ach so stimmt. Lucy bake ... bakes cakes.
 Ah that's right. Lucy bake ... bakes cakes.

In Example (6), one of the learners gives corrective feedback to the other, which leads to an immediate repair of the incorrect utterance. In this case, the design of the task results in drawing the learners' attention to the linguistic feature in focus (Doughty & Williams 1998a: 3).

- (6) C10 Okay, ehm ... Lucy like ice cream.
 C09 s!
 C10 Likes ice cream.

Finally, the dialogue in Example (7) illustrates that this kind of task-based work also results in negotiated interaction, in this case, negotiation of form. It is caused by Learner C10's use of the form 'joggings', which might reflect his regularisation

of an emergent grammatical structure (Pica 2005). This kind of interactive feedback has been shown to be effective for the repair of incorrect grammatical forms (see e.g. Nassaji & Fotos 2004:133). It ultimately leads to the use of a grammatically correct form of the participle (*gerund*) but in this context an inappropriate selection of the first verb. The fully correct form is finally supplied by C10's partner, C09.

- (7) C09 What does John do at Tuesday in the afternoon?
 C10 He ... joggings.
 C09 He is jogging.
 C10 He is jogging? Joggings? He is jogging. Joggings
 C09 He goes jogging.
 C10 Oh ja, Pech halt!
 Oh yes, that's bad luck!

The examples of task-based interaction above show that tasks with a developmentally moderated focus on form can lead to the use of targeted forms in many different ways and contexts. It can be seen that even primary school learners can deal with these kinds of tasks and reach the respective communicative goal, even though, as Pinter (2007: 189) reports, it is often assumed that they could not: "Teachers often feel that children at a low level of competence are generally unable to handle communication tasks and benefit from them in any way." Thus, tasks with a developmentally moderated focus on form can be used with learners of different age groups and at different levels of acquisition, provided that they are designed in ways that they are interesting and relevant for the respective target group and are based on vocabulary with which learners are familiar.

7. Summary

With regard to the three questions asked in the beginning, what to teach, when and how, it has been shown that by using communicative tasks with a developmentally moderated focus on form, it is possible to work out when specific linguistic features are learnable by particular learners and then to offer these learners productive learning opportunities. Tasks designed in this way "create contexts in which learners can experience what it means to communicate at different stages of their development", which, according to Ellis (2009: 230), is a main aim of task-based language teaching. Communicative tasks with a developmentally moderated focus on form can stimulate a kind of language use and interaction, which helps learners to experiment with and productively use forms they are developmentally ready

for. Thus, this approach has the potential to enhance and facilitate second language acquisition. Further research is needed in order to study the effects of the selective use of such tasks on acquisition. Since we know that learners bring a great potential to the classroom, using form-focused tasks that take developmental readiness into account could be a way of making use of it.

References

- Council of Europe (2001). *The Common European Framework of Reference for Languages: Learning, teaching, assessment*, Cambridge: Cambridge University Press.
- Di Biase, B. (2002). Focusing strategies in second language development: A classroom-based study of Italian L2 in primary school. In B. Di Biase (Ed.), *Developing a second language: Acquisition, processing and pedagogy of Arabic, Chinese, English, Italian, Japanese, Swedish* (pp. 95–120). Melbourne: Language Australia.
- Doughty, C., & Varela, E. (1998). Communicative focus on form. In Doughty & Williams (Eds.), pp. 114–138. Cambridge: Cambridge University Press.
- Doughty, C., & Williams, J. (Eds.). (1998). *Focus on form in second language acquisition*. Cambridge: Cambridge University Press.
- Doughty, C., & Williams, J. (1998a). Issues and terminology. In Doughty & Williams (Eds.), pp. 1–11.
- Doughty, C., & Williams, J. (1998b). Pedagogical choices in focus on form. In Doughty & Williams (Eds.), pp. 197–261.
- Ellis, R. (1989). Are classroom and naturalistic acquisition the same? A study of the classroom acquisition of German word order rules. *Studies in Second Language Acquisition*, 11(3), 305–328. doi:10.1017/S0272263100008159
- Ellis, R. (2003). *Task-based language teaching and learning*. Oxford: Oxford University Press.
- Ellis, R. (2005). *Instructed second language acquisition. A literature review. Report to the Ministry of Education*. Wellington: Ministry of Education.
- Ellis, R. (2009). Task-based language teaching: Sorting out the misunderstandings. *International Journal of Applied Linguistics*, 19(3), 221–246. doi:10.1111/j.1473-4192.2009.00231.x
- Engel, G., Groot-Wilken, B., & Thürmann, E. (Eds.). (2009). *Englisch in der Primarstufe - Chancen und Herausforderungen*. Berlin: Cornelsen.
- Kefler, J.-U. (2006). *Englischerwerb im Anfangsunterricht diagnostizieren. Linguistische Profilanalysen und der Übergang von der Primar- in die Sekundarstufe I* (Giessener Beiträge zur Fremdsprachendidaktik). Tübingen: Narr. doi:10.1075/palart.1
- Kefler, J.-U., Liebner, M., & Mansouri, F. (2011). Teaching. In Pienemann & Kefler (Eds.), pp. 150–156. Amsterdam: John Benjamins. doi:10.1075/palart.1.12tea
- Kefler, J.-U., & Pienemann, M. (2011). Research methodology: How do we know about developmental schedules? In Pienemann & Kefler (Eds.), p. 84–96. Amsterdam: John Benjamins. doi:10.1075/palart.1.07res
- Kefler, J.-U., & Plesser, A. (2011). *Teaching grammar*. Paderborn: Schöningh.
- Lenzing, A., & Roos, J. (2012). Die sprachliche Entwicklung und die Ausdrucksmöglichkeiten von Grundschülerinnen und Grundschulern im Englischunterricht. In M. Bär, A. Bonnet, H. Decke-Cornill, A. Grünewald, & A. Hu (Eds.), *Globalisierung – Migration – Fremdsprachenunterricht. Dokumentation zum 24. Kongress für Fremdsprachendidaktik der Deutschen Gesellschaft für Fremdsprachenforschung (DGFF) Hamburg* (pp. 207–220). Baltmannsweiler: Schneider Verlag Hohengehren.

- Lenzing, A. (2013). *The development of the grammatical system in early second language acquisition – The Multiple Constraints Hypothesis*. Amsterdam: John Benjamins.
doi:10.1075/palart.3
- Liebner, M., & Pienemann, M. (2011). Explaining learner variation. In Pienemann & Kefler (Eds.), pp. 64–74. Amsterdam: John Benjamins. doi:10.1075/palart.1.05exp
- Lightbown, P. M. (1998). The importance of timing in focus on form. In Doughty & Williams (Eds.), pp. 177–196.
- Lightbown, P. M., & Spada, N. (2013). *How languages are learned* (4th ed.). Oxford: Oxford University Press.
- Long, M. H. (1996). The role of the linguistic environment in second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 413–468). New York, NY: Academic Press.
- Long, M. H. (2000). Focus on form in task-based language teaching. In D. Richard & E. Shohamy (Eds.), *Language policy and pedagogy. Essays in honour of A. Ronald Walton Lambert* (pp. 179–192). Amsterdam: John Benjamins. doi:10.1075/z.96.110n
- Long, M. H. (2011). Methodological principles for language teaching. In C. Doughty & M. H. Long (Eds.), *Handbook of language teaching* (pp. 373–394). Oxford: Wiley-Blackwell.
- Long, M. H., & Robinson, P. (1998). Focus on form: Theory, research and practice. In Doughty & J. Williams (Eds.), pp. 15–41.
- Mackey, A. (1994). Targeting morpho-syntax in children's ESL: An empirical study of the use of interactive goal-based tasks. *Working Papers in Educational Linguistics*, 10(1), 67–91.
- Mackey, A. (1999). Input, interaction and language development. *Studies in Second Language Acquisition*, 21(4): 557–587. doi:10.1017/S0272263199004027
- Mansouri, F., & Duffy, L. (2005). The pedagogic effectiveness of developmental readiness in ESL grammar instruction. *Australian Review of Applied Linguistics*, 28(1), 81–99.
- Meisel, J. M., Clahsen, H., & Pienemann, M. (1981). On determining developmental sequences in natural second language acquisition. *Studies in Second Language Acquisition*, 3, 109–135. doi:10.1017/S0272263100004137
- Müller-Hartmann, A., & Schocker-von-Ditfurth, M. (2011). *Teaching English: Task-supported language learning*. Paderborn: Schöningh.
- Nassaji, H., & Fotos, S. (2004). Current developments in research on the teaching of grammar. *Annual Review of Applied Linguistics*, 24, 126–145. doi:10.1017/S0267190504000066
- Norris, J., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning*, 50, 417–528. doi:10.1111/0023-8333.00136
- Pica, T. (2005). Second language acquisition research and applied linguistics. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 263–280). Mahwah, NJ: Lawrence Erlbaum Associates.
- Pienemann, M. (1984). Psychological constraints on the teachability of language. *Studies in Second Language Acquisition*, 6, 186–214. doi:10.1017/S0272263100005015
- Pienemann, M. (1989). Is language teachable? Psycholinguistic experiments and hypotheses. *Applied Linguistics*, 10(1), 52–79. doi:10.1093/applin/10.1.52
- Pienemann, M. (1998). *Language processing and second language development: Processability Theory*. Amsterdam: John Benjamins. doi:10.1075/sibil.15
- Pienemann, M., & Kefler, J.-U. (Eds.). (2011). *Studying Processability Theory*. Amsterdam: John Benjamins. doi:10.1075/palart.1
- Pinter, A. (2007). Some benefits of peer–peer interaction: 10-year-old children practising with a communication task. *Language Teaching Research*, 11(2), 189–207.
doi:10.1177/1362168807074604

- Roos, J. (2007). *Spracherwerb und Sprachproduktion: Lernziele und Lernergebnisse im Englischunterricht der Grundschule* (Giessener Beiträge zur Fremdsprachendidaktik). Tübingen: Gunter Narr.
- Roos, J. (2012). Lernszenarien für die Entwicklung der sprachlichen Kompetenz im Englischunterricht der Grundschule. In H. Böttger & N. Schlüter (Eds.), *Fortschritte im frühen Fremdsprachenlernen. Ausgewählte Tagungsbeiträge Eichstätt 2011* (pp. 77–85). München: Domino.
- Samuda, V. (2001). Guiding relationships between form and meaning during task performance: The role of the teacher. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching pedagogic tasks: Second language learning, teaching, and testing* (pp. 119–134). Harlow: Longman.
- Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, 11(2), 129–58. doi:10.1093/applin/11.2.129
- Spada, N. (1997). Form-focussed instruction and second language acquisition: A review of classroom and laboratory research. *Language Teaching*, 30, 73–87. doi:10.1017/S0261444800012799
- Spada, N. (2010). Beyond form-focused instruction: Reflections on past, present and future research. *Language Teaching*, 44(2), 225–236. doi:10.1017/S0261444810000224
- Spada, N., & Lightbown, P. (2008). Form-focused instruction: isolated or integrated? *Tesol Quarterly*, 42(2), 181–207. doi:10.1002/j.1545-7249.2008.tb00115.x
- VanPatten, B., & Williams, J. (2007). Early theories in second language acquisition. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition* (pp. 17–35). Mahwah, NJ: Lawrence Erlbaum Associates.

Psychometric approaches to language testing and linguistic profiling – A complementary relationship?

Katharina Hagenfeld
University of Paderborn

The present study investigates as to whether and to what extent Linguistic Profiling can complement shortcomings of proficiency rating scales that are based on the Common European Framework of Reference (CEFR) (CoE 2001). In order to shed light on possible interfaces between the second language acquisition theory Processability Theory (PT) (Pienemann 1998, 2005) and the CEFR, learners were rated according to the CEFR and diagnosed with two linguistic profiling tools: Rapid Profile (Mackey, Pienemann, & Thornton 1991; Pienemann & Mackey 1993; Kefßler 2006, 2008) and Autoprofile (Lin 2012). The emergence criterion (Pienemann 1998; Pallotti 2007) as used in PT as the starting point to determine acquisition is highly predictive in nature and thus taken as the point of departure of an integration of PT into the CEFR. The results show correspondences between CEFR levels and PT stages and suggest a reexamination of early CEFR levels in terms of the complexity of operations beginning learners are assumed to manage.

1. Introduction

This study aims to (1) determine possible interfaces between psychometric rating scales that are based on the Common European Framework of Reference and the section of Linguistic Profiling (LP)¹ (Pienemann, Johnston Brindley 1988) which is situated within the psycholinguistic theory of second language acquisition (SLA) Processability Theory (Pienemann, 1998, 2005). It (2) makes inferences about a possible integration of the LP tools Rapid Profile (RP) (Mackey, Pienemann, & Thornton 1991; Pienemann & Mackey 1993; Kefßler 2006, 2008)

1. The term LP is adapted from its original association with: “The analysis of a person’s speech or writing, especially to assist in identifying or characterizing an individual or particular subgroup (cf. Oxford dictionary)”.

that works semi-automatically and/or the fully automatic diagnostic tool Autoprofiling (AP) (Lin 2012) into language proficiency ratings. An integration is assumed to enhance the objectivity, reliability and validity of rating procedures. In this way it contributes to bridging the gaps between language acquisition research and language testing. To make claims about integrating RP and/or AP into ratings, (3) both tools are compared in terms of reliability and time allotment for the procedure. It is then inferred as to which tool is more suitable for large scale assessment in combination with the CEFR.

The paper starts out with introducing the main ideas behind the psychometric approach to language testing that is taken up by the CEFR. Its weaknesses are presented in order to clarify why a complementation with a second language (L2) acquisition theory is important. The paper proceeds with elaborating on Processability Theory and its diagnostic tools. My point of departure in making a case for the integration of PT into ratings; i.e. the emergence criterion, is highlighted afterwards. The study, its aims and results are described in section four and discussed in part five in this paper. The conclusion offers a summary of the research findings and discussion.

2. Testing based on the CEFR – A psychometric approach

The development of the CEFR dates back to the 1970s in which a paradigm shift in language teaching and education evolved (cf. North 2007). As opposed to traditional; teacher centered, teaching methods such as the Grammar-translation method,² more learner centered and communicative approaches such as Communicative Language Teaching (CLT) and Task-based Language Teaching (TBLT) arose. The CEFR claims to reflect these changes in describing an action-oriented approach³ to language use (under which language acquisition is subsumed, cf. CoE 2001:21) that hypothesizes the development of L2 proficiency to be based on the usage of communication and communicative strategies and activities (CoE 2001:9). The emphasis on communication and communicative acts is reflected in the framework that seeks to provide “a common basis for the elaboration of language syllabuses, curriculum guidelines, examinations, textbooks, etc. across Europe” (CoE 2001:1). The CEFR itself thus does not mean to test language

2. For a discussion on the history of language teaching and traditional teaching methods, see Kefler & Plesser (2011).

3. Action-oriented approach regards the notion in the CEFR that language learners and users are social agents and that every communicative act is socially founded (cf. CoE 2011:21).

proficiency as such but to provide a basis on which tests can be designed and administered across European member states. The approach to language assessment that is suggested in the framework is proficiency testing with rating scales that originate in psychometric studies.

Psychometric language testing evolved out of the scientific field of psychology in order to provide objective measures for subjective items such as personality traits, attitudes and academic achievements (cf. Michell 1999; Kaplan & Saccuzzo 2010). The assumed objectivity is achieved through the use of questionnaires and scales (cf. Stevens 1946) that describe an item, such as a personality trait, and can thus be matched to the perceived reality of the person to be tested. In the case of the CEFR, the matter to be tested is language proficiency. In using an action-oriented approach, the European framework defines language proficiency to be based on a number of competences which “[...] are the sum of knowledge, skills and characteristics that allow a person to perform actions.” (CoE 2006: 9). There are general competences which are “[...] not specific to language, but which are called upon for actions of all kinds, including language activities.” (CoE 2006: 9). The competencies are subdivided into several language skills. These communicative skills are described in the global scale, Figure 1, that is “arranged in three bands – A1 and A2 (basic user), B1 and B2 (independent user), C1 and C2 (proficient user)” (Little 2008: 4). Each level provides descriptors as to the skills that need to be attained to reach a certain level.

The global scale is supposed to provide points of orientation for teachers and curriculum planers (cf. CoE 2001). Additionally, the Council of Europe caters scales for communicative tasks at different levels such as oral/written production and comprehension as well as self-assessment scales. Apart from the broad benefits the CEFR was able to manifest, such as encouraging a basis for a cooperation between educational institutions all over Europe, formulating a common ground of criteria for qualifications in the area of language and providing access to cultural manifestations (CoE 2001: 17), it has to face extensive critique when it comes to being the basis for language testing. The following section illustrates major points of critique but raises no claim to completeness; it rather provides a brief overview of points of critique relevant for this study.

2.1 Critique as regards psychometric testing and the CEFR

Psychometric testing itself has been subject to extensive critique for a number of reasons out of which the following four points will be further discussed: (1) rating scales operate within human limitations; (2) they do not measure directly but through introspection, i.e. post factually, which implies a threat to objectivity. In relation to *language* testing, the following points are criticized: (3) the concept

Proficient	C2	Can understand with ease virtually everything heard or read. Can summarise information from different spoken and written sources, reconstructing arguments and accounts in a coherent presentation. Can express him/herself spontaneously, very fluently and precisely, differentiating finer shades of meaning even in more complex situations.
User	C1	Can understand a wide range of demanding, longer texts, and recognise implicit meaning. Can express him/herself fluently and spontaneously without much obvious searching for expressions. Can use language flexibly and effectively for social, academic and professional purposes. Can produce clear, well-structured, detailed text on complex subjects, showing controlled use of organisational patterns, connectors and cohesive devices.
Independent	B2	Can understand the main ideas of complex text on both concrete and abstract topics, including technical discussions in his/her field of specialisation. Can interact with a degree of fluency and spontaneity that makes regular interaction with native speakers quite possible without strain for either party. Can produce clear, detailed text on a wide range of subjects and explain a viewpoint on a topical issue giving the advantages and disadvantages of various options.
User	B1	Can understand the main points of clear standard input on familiar matters regularly encountered in work, school, leisure, etc. Can deal with most situations likely to arise whilst travelling in an area where the language is spoken. Can produce simple connected text on topics which are familiar or of personal interest. Can describe experiences and events, dreams, hopes & ambitions and briefly give reasons and explanations for opinions and plans.
Basic	A2	Can understand sentences and frequently used expressions related to areas of most immediate relevance (e.g. very basic personal and family information, shopping, local geography, employment). Can communicate in simple and routine tasks requiring a simple and direct exchange of information on familiar and routine matters. Can describe in simple terms aspects of his/her background, immediate environment and matters in areas of immediate need.
User	A1	Can understand and use familiar everyday expressions and very basic phrases aimed at the satisfaction of needs of a concrete type. Can introduce him/herself and others and can ask and answer questions about personal details such as where he/she lives, people he/she knows and things he/she has. Can interact in a simple way provided the other person talks slowly and clearly and is prepared to help.

Figure 1. Global Scale, taken from the Manual for LTD (CoE 2012)

language that is to be measured is not clearly defined to be readily operationalized for testing purposes. As for the CEFR, much work has been put into the careful formulation of descriptor items based on sociological and philosophical ideas but (4) it still lacks a comprehensive theory of language and its acquisition.

With regard to (1) trained raters use descriptive scales in order to, in the case of language testing, assign learners a language proficiency level. Much work was put into the development of assessment criteria grids that help the analyst with their rating. However, rating scales work only as well as the person who uses them. Biases

were reported on cultural levels (Rohrman 2007: 1), ambiguous interpretation of descriptor items, how harsh or lenient raters score with regard to overall performance, traits (Schaefer 2008) and subject groups (Wigglesworth 1993) as well as accent familiarity (Winke, Gass, & Myford 2013) among others. Irrespective of how strong or weak these factors may influence ratings, they cannot be fully eradicated as raters are sensitive to one factor or another as no person is fully objective.⁴

(2) Indirect measures are generally perceived as being less concrete than direct measurements. They are usually based on (self-) reports or questionnaires about a behavior, skill or else. The crux here is that the item to be measured is assessed through retro- or introspection. The question remains as to whether it can be determined that what the rater perceives reflects the reality. Thus, the dependence of the test result on the raters' opinions, judgments and beliefs forms a major drawback in terms of objectivity.

As with psychological variables, (3) language itself is not a concept that is easily defined and operationalized. The matter to be tested is language. Language is built of sounds, intonation, stress, morphemes, words, and arrangements of words having meanings that are linguistic and cultural.[...] They are integrated in the total skills of speaking, listening, reading and writing [...] all of which do not advance evenly. (Lado 1961:25). The broad scope of the concept language makes it hard to determine and test all of its properties.

Since there is, as of yet, no universally accepted and operationalized definition of language proficiency (cf. Pienemann & Kessler 2007:247) the development of respective tests relies on the definition of the test administrator. With regard to the CEFR (4), language is defined as action-oriented in which learners are seen as subjects who operate in varying social contexts and who have to fulfil varying social activities (CoE 2001:21). In order to be able to act creatively with language, they thus have to acquire certain communicative competences and strategies (CoE 2001:21). Harsch (2005:26) criticizes the vague definition of the term language as well as the CEFRs' equalization of the terms language use and language acquisition (Ibid., p. 65.). How can we test something without knowing what it is that we want to test? Much effort has been put into the development of theories of language use and language acquisition that are hardly touched upon in the CEFR. This means that (4) a comprehensive theory or approach behind the CEFR cannot be found. However, it is based on principles that go back to the philosopher Dell Hymes (1974) who hypothesized the development of an individual as being promoted by the acquisition of different competencies while completing everyday activities

4. However, Wigglesworth (1993) found that raters tend to react to feedback and their willingness to change their behavior to achieve more objective scores.

(c.f. Dell Hymes 1974). The broad, action-oriented scope of the term language that the CEFR seeks to cover adds to the difficulty in finding an appropriate theoretical foundation. Thus the CEFR maintains quite a low profile in this regard. Since this influential framework has such a far-reaching influence, it is claimed that it is important to constantly reflect on it and enhance it by including current trends and findings in (second) language acquisition research. Having outlined major points of critique as regards the CEFR, the next chapter focuses on language testing based on a theory of language development, i.e. Processability Theory.

3. Assessing interlanguage development with Rapid Profile and Autoprofiling

This chapter briefly introduces Processability Theory and its diagnostic tools Rapid profile and Autoprofiling. It will end with listing advantages of these testing procedures concerning a possible enhancement of quality testing criteria in the language proficiency ratings.

Processability Theory (PT) (Pienemann 1998, 2005) is a psycholinguistic theory of L2 development that predicts a universal developmental hierarchy for the acquisition of a second language. The hierarchy is spelled out based on the architecture of the human language processor as modeled by Levelt's blueprint for the speaker (1989).⁵ PT's core assumption is that a linguistic structure can only be produced and consequently acquired if the current state of the language processor is capable of processing the respective linguistic form (Pienemann 2007: 137). Initial psychological constraints account for a cumulative and successive acquisition process that is implicationally related. An implicational order infers that a later structure and procedure "implies the presence of an earlier structure" (Pienemann 2011: 51; Pienemann & Keßler 2011). Two key mechanisms are crucial in this regard: (a) feature unification and (b) specific mapping processes as modeled in Lexical-Functional Grammar (Bresnan 2001).⁶ Feature unification accounts for the matching of grammatical features as produced by Levelt's

5. Due to the scope of this paper, a detailed description of Levelt's model of message generation will not be given. For further information, please see (1989).

6. Lexical Functional Grammar is a formal theory of grammar that assumes three levels of linguistic representation to be linked by mapping principles. Argument structure contains the meaning to be expressed (verb and its arguments) that is to be mapped onto the surface form as represented in constituent structure. Functional structure in which grammatical features are encoded links argument and constituent structure. Please see Bresnan (2001) and Lenzing (2013) for more information.

processing components (Levelt 1989:9). LFG's mapping principles allow for the spelling out of the developmental hierarchy for a variety of typologically diverse languages such as Chinese, English, Swedish, etc. PT therefore, does not use the term proficiency in L2 but development. L2 development is defined as those grammatical features that are captured in the hierarchy. The hierarchy for English as a second language is spelled out as follows:

Stage	Processing procedures	Phenomena	Examples
6	Subordinate clause - procedure	Cancel Aux-2nd	I wonder <i>what he wants</i> .
5	S-procedure	Neg/Aux-2nd-? Aux-2nd-? 3sg-s	Why <i>didn't</i> you tell me? Why <i>can't</i> she come? Why <i>did</i> she eat that? What <i>will</i> you do? Peter likes bananas.
4	VP- procedure	Copula S (x) Wh-copula S (x) V-particle	<i>Is</i> she at home? <i>Where</i> is she? Turn it <i>off!</i>
3	Phrasal procedure	Do-SV(O)-? Aux SV(O)-? Wh-SV(O)-? Adverb-First Poss (Pronoun) Object (Pronoun)	<i>Do</i> he live here? <i>Can</i> I go home? <i>Where</i> she went? <i>What</i> you want? <i>Today</i> he stay here. I show you <i>my</i> garden. This is <i>your</i> pencil. Mary called <i>him</i> .
2	Category procedure	S neg V(O) SVO SVO-Question -ed -ing Plural -s (Noun) Poss -s (Noun)	<i>Me no</i> live here. / I don't live here. Me live here. You live here? John <i>played</i> . Jane <i>going</i> . I like cats. Pat's cat is fat.
1	Word / lemma access	Words Formulae	Hello, Five Dock, Central How are you? Where is X? What's your name?

Figure 2. PT hierarchy for English as a L2, adapted and modified from Pienemann (2005:24)

Figure 2 depicts the universal processing procedures at the left hand side along the structural linguistic realization of these exemplified with illustrations. Stage one represents the lowest stage at which words and formulaic utterances can be retrieved from the mental lexicon (as modelled in Levelt (1989)). The higher the stage, the more productive the learner becomes in the target language.

In order to capture and use the power of PT to predict the course of L2 development, diagnostic tools to determine the current stage of acquisition of L2 learners were developed. A benefit of knowing the current stage of learner development is to enable language instructors to provide targeted instruction and to pick the learners up where they are.

Linguistic Profiling is based on theoretical work by Crystal et al. (1982) and Clahsen (1985) and follows the profiling approach by Crystal, Fletcher, and Garman (1976) in the domain of language disorders⁷ and its adoption and modification for German L2 acquisition by Clahsen (1985). For a profile analysis an interview is conducted, a full transcription of the interview is done on which a careful analysis of the sample is based (Pienemann & Mackey 1993:24). The profile approach as used with LARSP is in line with the view of language assessment taken by Pienemann (1998) as it is

(a) descriptive, (b) developmental, and (c) interactive. The first refers to the descriptive categories provided by the procedure; the second, to the developmental schedule of these categories; and the third, to the method of data collection: the spontaneous speech gathered in unstructured conversations. [...] Descriptive criteria are objective; developmental criteria are psychologically plausible, and interactive criteria are based on natural language use.

Pienemann, Johnston, & Brindley (1988:231)

For a language profile, natural oral speech data is elicited and scrutinized using distributional analysis. Distributional analysis as used by Pienemann (1998:139) allows for determining “which contexts or even which lexical items are related to which particular interlanguage rules.” This way, idiosyncratic and formulaic use of the target language can be ruled out (please see the following section on the emergence criterion for further explanation). Since a careful and fine-grained distributional is quite time-consuming and hardly feasible in ESL contexts, a rapid version for the allocation of interlanguage development was established, i.e. Rapid Profile (RP). *RP* (Mackey, Pienemann, & Thornton 1991; Pienemann & Mackey 1993; Keßler 2006, 2008) is a computer-assisted screening procedure operated by trained linguistic profilers (cf. Keßler 2006). It is a short-hand version of the original linguistic profile. In *RP* the profiler uses communicative tasks that trigger the production of a specific linguistic structure found in the Processability hierarchy. The tasks focus only on those linguistic items that are crucial for determining the

7. Language Assessment Remediation and Screening Procedure (LARSP) is a screening procedure to allocate learner language in terms of grammatical disability. It has been widely used up to date, mainly by speech therapists and language researchers. For more information, see Crystal, Fletcher, & Garman (1976), Crystal (1982).

developmental stage of a learner (cf. Pienemann & Mackey 1993:25). They usually contain an information gap to make the subjects produce specific linguistic structures without noticing them. The 3rd person singular-s, for example is elicited through a habitual action task that shows pictures of the daily routine of a fictive person. In order to gain a dense data set, at least three different tasks aiming at different structures in the Processability hierarchy need to be used. During the production phase of the learner, the profiler uses the computer interface to check those buttons that relate to the structure produced. Figure 3 shows the RP interface with the boxes for each structure in the hierarchy.

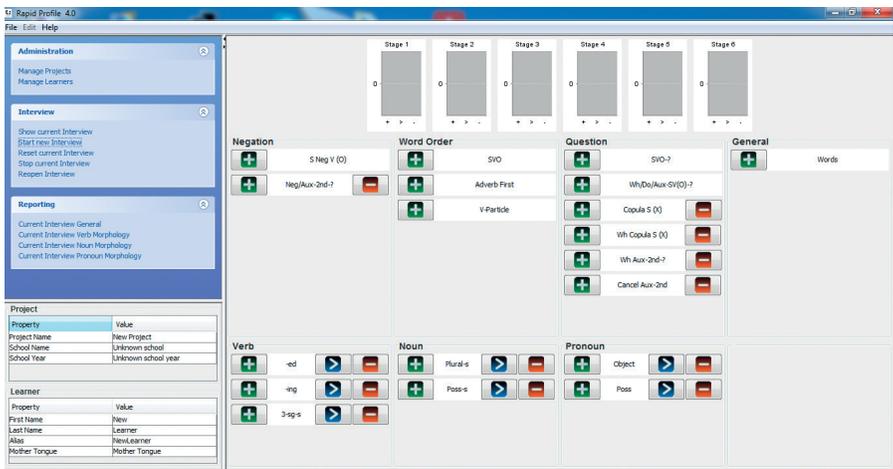


Figure 3. Rapid Profile 4.0 user interface

The structures are subdivided into syntactic phenomena on the top and morphological phenomena on the bottom of the interface. If the learner produces a verb in an obligatory context along with a morphological feature such as the past-ed, the profiler clicks on the plus under the headline verb. Should the learner fail to attach a past-ed in an obligatory context, the profiler would check the minus-box. The program computes the developmental stage in checking the data typed in against standard learner language according to the *emergence criterion* (see the next section for further information). Rapid Profile gives detailed feedback not only on the developmental stage but morphology and syntax. The very recent development of *Autoprofiling (AP)* (Pienemann, Lin, & Chung 2009; Lin 2012) further simplifies this procedure by working analogous to RP. AP is an online screening procedure that operates fully automatically. With AP, the learner simply types in his/her answers to the tasks into an input field. The interlanguage sample is calculated in comparison to a small corpus

that is embedded in the program (Lin 2012). The sample is conducted under a time constraint that rules out the usage of declarative monitoring.⁸ The program works similar to a common RP analysis but is accessible from everywhere and at anytime. As opposed to RP, AP works with written input. This however should not have any influence on the developmental level of the learners since the mode-steadiness hypothesis (Plesser 2008) predicts that interlanguage development remains within the concept of hypothesis space even across mode-barrier boundaries. Håkansson and Norrby (2006) underline Plesser's findings in comparing written and spoken L2 Swedish. Their learners followed the PT hierarchy on both modes with a tendency for learners to be one level ahead in written production.

Advantages of Rapid Profile lie in the computer-assisted nature of the program that compares standard patterns of development with a learner's interlanguage sample (Keßler 2006). Thus, the program scores high in objectivity. Trained profilers are able to elicit a profile with high inter-rater-reliability (Keßler 2006: 241). This is why the use of RP allows for accommodating reliable and valid results in only fifteen minutes (Keßler 2006). In his study, Keßler (2006) tested as to whether fifteen minutes were sufficient enough in order to elicit a dense data set. His results showed that "[...] the data elicitation took an average of 12.5 minutes and ranged between seven and 17 minutes" (Keßler & Plesser 2011: 214) with sufficient data density.

While Rapid Profile establishes high standards for language testing in terms of rapidness and reliability, a disadvantage is the profiler-dependent usage of the program. Autoprofiling by Lin (2012) has the potential to overcome this limitation since it operates fully automatically. Lin (2012) showed that there is 99.0% accordance of AP results compared to RP results. To recapitulate, LP scores sharply high not only in measurement and testing criteria as defined by Bachman (1990, 2004), Rasinger (2008) or Neuendorf (2002) but includes a very detailed interlanguage and grammatical feedback. Thus LP is able to provide hands-on feedback to the language learner with clear indications as to what s/he is able to do at this specific point in her/his development as well as what is learnable next.⁹ This predictive power that underlies the results of LP further enables teachers to

8. The time constraint is embedded to avoid that a learner goes back to her/his written work and changes it terms of style and accuracy as this would lead to a distortion of the profile. PT assumes procedural knowledge to be more important in the production of a second language than declarative knowledge. For further information, please see Plesser (2008), Ellis (2005, 2007).

9. This is one of the criteria that Brindley (1998: 117) considers to be crucial in language testing.

internally differentiate their students according to their developmental levels and give respective instruction. The following section further elaborates on the predictive power of PT.

3.1 The emergence criterion in Rapid Profile and Autoprofiling

The emergence criterion (EC) (Meisel, Clahsen, & Pienemann 1981; Pienemann 1998; Pallotti 2007) assumes that as soon as a grammatical structure appears in the interlanguage of a learner, the structure is assumed to be acquired (Kefler 2006: 147; Pienemann & Kefler 2011). Thus, the production of a linguistic structure defines “[...] the beginning of an acquisition process, and focusing on the start of this process will allow the researcher to reveal more about the rest of this process.” (Pienemann 1998: 138). Thus, RP and AP need three incidences of syntactical and morphological and lexical variation to assume a structure to be acquired. Morphological and lexical variation is exemplified in Figure 4 with the help of three verbs and three morphemes below.

Lexical variation	Morphological variation
go	-ing
+	+
play	-ed
+	+
talk	-(e)s

Figure 4. Illustration of emergence criterion

Thus, to predict that the third-person singular-s is de facto acquired, it would need to be attached to all verbs above and the learner would have to use the different inflectional endings as well. This way, a mere storage of stem and affix as a chunk in the mental lexicon (as modelled in Levelt 1989) can be ruled out as the learner has to use both, ending and verb, creatively. The emergence criterion allows one to pinpoint the acquisition of an underlying interlanguage structure in a direct manner. Using the EC as the point of acquisition, no third party has to judge or rate whether a feature might have been attained. Unlike with rating scales, the usage of introspective or retrospective means (such as questionnaires or thinking aloud protocols) on the sides of learner is not necessary either.

In making a case for the integration of linguistic profiling into the CEFR, the emergence criterion as an objective means to indicate language development

strengthens the validity of PT and its assessment tools and is thus considered crucial in this context.

In the following section, the study and its components along the results are outlined.

4. The study

This cross-sectional study is a pilot study which was conducted in order to make inferences about a possible integration of linguistic profiling into the Framework of References for Languages. One has to note however, that PT takes a modular approach to LA whereas the CEFR covers language and its acquisition holistically.

4.1 Aims and research questions

The overall goal of this study is to investigate whether LP can be used as a complementary assessment tool to the CEFR in order to provide a more objective, reliable and accurate feedback for the testees. Thus, my aims are the following: (a) In order to hypothesize an integration of linguistic profiling into the CEFR, it must be ascertained that there are correspondences between the CEFR levels and the developmental stages as predicted by PT. This study therefore takes another look at the findings by Lenzing and Plesser (2010) who piloted the quest for correspondences. In a pilot study, they found that PT stage three relates to CEFR level A1, stage four to A1 and B1, stage five to B1, B2 and C1 and stage six to C1 as can be seen in Figure 5.

Rapid Profile	CEFR Level
Stage 1	Below A1
Stage 2	Below A1 and A1
Stage 3	A1
Stage 4	A1, B2
Stage 5	B1, B2, C1
Stage 6	C1

Figure 5. CEFR and Rapid Profile correspondences, Lenzing & Plesser (2010)

The present study reconsiders their findings with further data. Once correspondences are laid out, it is believed that a combination of LP based on PT and the CEFR can cover many aspects of language proficiency and development with

enhanced results in objective feedback and reliability. RP and AP's beneficial backwash then allows teachers to provide learners with materials based on the predictions of PT to help them progress in their interlanguage development.

Another aim of this paper is to examine the relationship between Rapid Profile and Autoprofile. As mentioned in Section 3, Autoprofile is assumed to be (b) more feasible in large-scale assessment settings than Rapid Profile is due to the fully automatic nature of the program. One has to however, go a step back and examine whether Autoprofile shares RP's benefits in terms of (c) reliable diagnostic outcome and whether AP may even exceed RP in the time span needed for the assessment. A faster assessment is generally seen as more feasible. Without making sure that the latest addition to the PT formula is as reliable as its predecessor, making claims about its integration into the CEFR is superfluous. This is why RP and AP feedback will be compared in terms of the developmental stage and the time allotment both programs compute.

These aims generate the following research questions:

- a. Are there are correspondences between CEFR levels and PT stages?
- b. Is AP more feasible than RP in rating settings due to its profiler-independence and rapidness?
- c. Are RP and AP equally reliable?

I hypothesize that there is indeed correspondence between CEFR levels and PT stages as indicated by the findings by Lenzen & Plesser (2010). Due to the nature of AP, I further hypothesize that AP scores higher than RP in time allotment and infer that both programs are equally reliable in terms of feedback. In order to test these claims, the data were elicited as follows.

4.2 Data and methodology

For the study, speech samples of nine university students were collected out of which three were male and six female. The students attended English courses at different CEFR levels from B1 to C2. Their professions differ widely, ranging from mechanical engineers (three male students), sport science students (two female students), business students (two female students) to a teacher trainee (one female student). A biodata-questionnaire elicited the reasons why the participants took part in English-courses. 90% of them wrote they wanted to *refresh* their English and 10% take the course as a preparation for their future occupation. Prior to the courses, the students either took the Oxford Placement Test (OPT) online or they were rated by trained raters. The OPT shows high correspondence between the score the participants achieved and the CEFR levels. The OPT scale and the according recommended courses can be viewed from the appendix in this volume. Every learner participating in this study was thus rated according to the CEFR

and profiled with Rapid Profile as well as Autoprofiling in order to compare their CEFR level and PT stage.

Once the participants completed the OPT, they attended the respective language course. These courses run on a weekly basis. At the end of the course, a test following CEFR criteria is taken which shows whether the students fulfil the requirements. In working with these learners, it was assured that they were all sufficiently rated according to CEFR standards.

The RP profile was conducted with two (and three participants at level B1) at each CEFR level performing communicative tasks in pairs. The participants were audio-recorded and the recordings were transcribed. The RP analysis was conducted by a trained linguistic profiler and checked against the transcriptions. After the RP profile, the students were briefly introduced to the AP interface. They were given time to familiarize themselves with the program and had the opportunity to ask technical questions. Since AP operates fully online, the software can be used anywhere and anytime without a profiler being present. Thus, the students were not able to ask questions during the procedure. Both, RP and AP work with the same task design that only differs in subject matter. The communicative tasks that were used provide natural obligatory contexts for specific linguistic structures to be found in the PT hierarchy and are in line with the standards set by Mackey, Pienemann, & Thornton (1991), Pienemann & Mackey (1993).

Rapid Profile				Autoprofile		
Task Type	Habitual Action	Spot the difference	Interview	Picture description	Habitual Action	Interviews
	Describe the daily routine of Mr. and Mrs. Lee.	These are two pictures, they look similar but they are not the same. Ask questions to find out about the differences.	I am a Martian and you are a reporter. You can ask me whatever you want to know about me.	Describe the two pictures.	What does your mother, father, sister, brother or friend do every day?	You can ask these boys whatever you like.
Structural Outcome	SVO, adverbials, 3rd-ps-sg-s	Do/Aux-fronting, WH-cop-?, Wh-Aux-2nd -?	Do/Aux-fronting, WH-cop-?, Wh-Aux-2nd -?	SVO, adverbials, 3rd-ps-sg-s	SVO, adverbials, 3rd-ps-sg-s	Do/Aux-fronting, WH-cop-?, Wh-Aux-2nd -?

Figure 6. Tasks and the linguistic structures they trigger¹⁰

10. This is a summary based on work by Meisel, Clahsen, & Pienemann (1981); Mackey, Pienemann, & Thornton (1991); Keßler (2006), (2008); Lenzing (2010); Plesser (2011).

Figure 6 shows the tasks used in RP and AP and those linguistic forms that are triggered due to the task design. With RP, three tasks are used in which the profiler has to guide and moderate the conversation in order to gain enough data. In AP, four tasks are administered in which two aim at declarative sentences (picture description and habitual action task) whereas the two interviews trigger interrogatives. In this way, a most holistic profile in terms of syntactical and morphological features that are captured in PT is established.

The time it took the participants to complete the AP analysis was determined and compared to the length of the audio-recording for RP. For each subject, the CEFR level and developmental stage elicited with RP and AP were recorded. The results are as follows.

4.3 Results

Due to the limited number of participants, the project did not allow for the inclusion of learners at level A1. The lowest CEFR level captured here is thus represented by intermediate learners F01 to F03 at level B1 (Figure 7).

Learner	RP	AP	CEFR
F01	5	5	B1
F02	5	5	B1
F03	5	5	B1
F04	5	2	B2
F05	5	5	B2
F06	5	5	C1
F07	5	5	C1
F08	5	5	C2
F09	5	5	C2

Figure 7. CEFR levels and PT stages for RP and AP

With only one exception, i.e. learner F04, all learners who are at different CEFR levels were assessed to be at PT stage 5. Stage five is rather high in the PT hierarchy since feature unification takes place across phrase level allowing for subject-verb agreement to take place (see section three). As mentioned before, learner F04 at CEFR level B2 shows a different result in the assessment with RP and AP as Auto-profile generated developmental stage two for her. This result is rather astounding as stage two marks an early phase in L2 acquisition at which the grammatical

category, i.e. lexical morphemes, such as the past-ed in *goed can be unified. The reason as to the stage gap that learner F04 shows might lie in an insufficient handling of Autoprofiling. This will be discussed in the following section. It is a rather long way¹¹ from stage two to stage five in which the cognitive effort needed to process the information increases.

Comparing the present results with the findings by Lenzing and Plesser (2010), it can be seen that the results coincide for learners at PT level 5. To recapitulate, they found that PT stage two and three relate to level A1, PT stage four to A1 and B1, PT stage five to B1, B2 and C1 and PT stage six to C1. PT stage one and two are not captured by CEFR in that learners at these stages usually produce unanalyzed chunks of words that are memorized and stored and retrieved from the lexicon. The CEFR thus seems not to account for learners at the *very* beginning of the acquisition process. A general result that can be retained from both Lenzing and Plesser's as well as these findings is that their hypothesis "learners who have reached B1 (CEFR) are generally assessed stage 5 or higher" can be underlined. All learners are at PT stage 5 whereas their CEFR levels range from B1 to C2. The results support the findings by Lenzing and Plesser (2010) for higher stages.

Going back to research question (c), the reliability aspect in RP and AP, Figure 7 shows that except for subject F04, RP and AP give the exact same developmental stage which indicates a high reliability. Bearing subject F04 as an exception in mind, the reliability coefficient (r) calculates 0.944. At 94%, reliability can be considered to be profoundly high. Anecdotal evidence suggests that the exception with learner F04 is due to problems with handling the program. This issue will be discussed in detail in the next section of this paper.

Turning to hypothesis (b), the feasibility aspect in AP, it can be noted that except for the short introduction, indeed no profiler was needed for the assessment. However, it was also found that all learners were negatively affected by the narrow time constraint embedded in AP. This will also be further discussed in the next section.

Along with the feasibility aspect goes the time allotment for RP and AP assessments. Please note that one RP sample has to be excluded here since learners F01 and F02 participated in a dyad but performed the AP analyses separate from each other. As a consequence, the exact time frame for each participant cannot be estimated for RP.

11. This is not to say that it will ultimately take a long time to progress in the hierarchy since the rate of acquisition is unique in every individual (cf. Pienemann 1998, 2005).

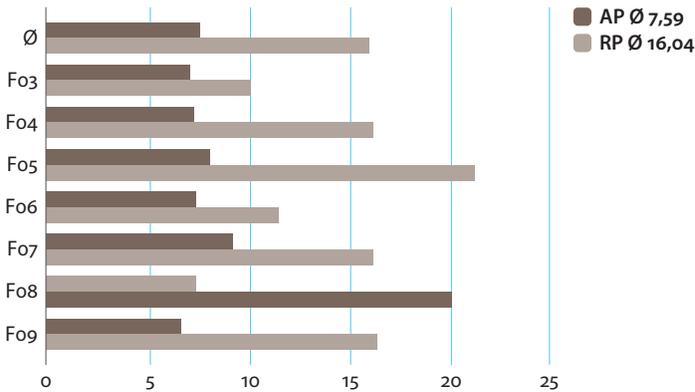


Figure 8. Time allotment for an AP and RP analysis in minutes

The horizontal axis in Figure 8 shows the time allotment in minutes whereas the vertical axis depicts the different learners that participated. At the top of the diagram, the average length of the interviews is given. The upper bar represents Auto-Profiling whereas the lower one stands for Rapid Profile. The arithmetic mean for time allotment with RP is 16 minutes and four milliseconds. For Auto-Profiling the arithmetic mean reveals seven minutes and fifty-nine seconds. It can thus be stated that AP is roughly twice as fast as RP is. The discursive nature and consequently the level of contribution of the profiler in this time span has to be considered. The results will be discussed in the following section with recourse to the handling errors that occurred with AP.

4.4 Discussion

It might be useful to briefly summarize the findings at this point. (A) Correspondences between CEFR levels and PT stages at higher stages were found. (B) RP and AP give the same results in terms of the developmental stage, although AP is vulnerable to handling errors that potentially distort the results. Because of the profiler-independence and its non-dialogical nature, AP is roughly twice as fast as RP in assessing the developmental stage of a learner. It is assumed these facts make (c) AP more feasible in large-scale rating settings as is the case with the CEFR but that AP in its current state needs a more fine-grained adjustment of the user interface.

As for result (a), we can hold that beginning learners are defined differently in the CEFR and PT not only in terms of their competence/development but in the overall concept of beginning learners as far as the global scale is concerned. Since PT, however, takes a modular approach to language acquisition that focuses on the prediction and explanation of grammatical development, the descriptors of the CEFR for grammatical accuracy might give more detailed insight into where

relations to PT can be found. These descriptors are presented in Figure 11 below. Grammatical accuracy in the CEFR is subsumed under linguistic communicative language competence within the area of control.¹²

Control

GRAMMATICAL ACCURACY

C2	<i>Maintains consistent grammatical control of complex language, even while attention is otherwise engaged (e.g. in forward planning, in monitoring others' reactions).</i>
C1	<i>Consistently maintains a high degree of grammatical accuracy; errors are rare and difficult to spot.</i>
B2	<i>Good grammatical control. Occasional "slips" or non-systematic errors and minor flaws in sentence structure may still occur, but they are rare and can often be corrected in retrospect.</i>
	<i>Shows a relatively high degree of grammatical control. Does not make mistakes which lead to misunderstanding.</i>
B1	<i>Communicates with reasonable accuracy in familiar contexts; generally good control though with noticeable mother tongue influence. Errors occur, but it is clear what he/she is trying to express.</i>
	<i>Uses reasonably accurately a repertoire of frequently used "routines" and patterns associated with more predictable situations.</i>
A2	<i>Uses some simple structures correctly, but still systematically makes basic mistakes- for example tends to mix up tenses and forget to mark agreement; nevertheless, it is usually clear what he/she is trying to say.</i>
A1	<i>Shows only limited control of a few simple grammatical structures and sentence patterns in a learnt repertoire.</i>

Figure 9. CEFR basic user at level A1, (CoE 2001: 114)

The illustrative scale for level A1 puts forward that in terms of grammatical accuracy the learner can handle limited simple grammatical structures and sentence patterns that were learnt. This converges with PT's prediction that at stage one, the learner uses formulaic sequences; i.e. unanalyzed chunks which are merely stored in and retrieved from the mental lexicon. However, Lenzing and Plessner's findings (Figure 5) reveal that learners at PT stage 1 learners are actually considered to be *below* CEFR level A1 in their study. Level A2 of the European framework states that some simple structures are used correctly but that mistakes are common. Those are stated however not to inhibit from getting the message across. Conceptually, Level A1 and PT stage one as well as level A2 and PT stage two seem

12. In The CEFR, control is defined as "Illustrative scales are available for the range of vocabulary knowledge, and the ability to control that knowledge (CoE 2001: 111).

to overlap since PT stage two is categorized by assigning the grammatical category and attaching lexical morphemes (category procedure) resulting in the production of phrases like: **Me live here, John played*, and interrogatives such as: **You live here?* As can be seen in Figure 12 below.

CEFR Level	Descriptors for grammatical accuracy	PT stage	Morpho-syntactic operation
A1	Shows only limited control of a few simple structures and sentence patterns in a learnt repertoire.	1	Words, formulae
A2	Uses some simple structures correctly, but still systematically makes basic mistakes – [...] forget to mark agreement [...].	2	Me no live here. / I don't live here. Me live here. You live here? John played. Jane going.

Figure 10. comparison waystage CEFR descriptors and initial PT stages

In PT terms, the learner is thus able to cover basic conversational acts with some errors in the same manner the CEFR assumes. Especially the descriptor about a lack of agreement marking in the CEFR on level A2 reflect the predictions by PT as this complex procedure is only able to be processed and produced at stage 5.

As opposed to the apparent conceptual overlap in PT stage 1 and CEFR level A1 as well as PT stage 2 and CEFR level A2, the findings by Lenzing & Plesser's empirical study reveal that level A1 relates to PT stage 2 and 3 and (for one learner even to) four. In PT terms this means that for stage three, the learner is able to unify information at the phrasal level, resulting in utterances such as **Do he live here?*, **Where she went?*, **What you want?*, **Today he stay here.*, I show you *my* garden., This is *your* pencil. Mary called *him*. (Pienemann 2005: 24). Utterances at PT stage three are much more productive and creative than "patterns in a learnt repertoire" (CoE 2001: 114). As mentioned before, it has to be borne in mind that this study was piloting the quest for relations between the CEFR and PT and thus had a small number of participants. This way, results are to be seen as tentative and hardly generalizable. Lenzing & Plesser's efforts can give hints though as to issues with the rating procedure itself. Analogous to critique mentioned in Chapter 2.1, the question remains whether the descriptors in A1 and A2 are explicit enough for a rating. It seems as if the subjectivity issues depicted in Section 2.1 are taking their toll on this study as well. Issues in rater biases and explicitness of descriptor items in rating procedures in language testing thus need further investigation.

Interestingly, the descriptors for level A1 on the global scale depicted in Figure 1 and the following elaboration by North (2007:7) indeed reflect the discursiveness that can be found at PT stages 2 and 3 as identified by Lenzing and Plesser (2010).

In making recourse to Wilkins (1978) and the Swiss Project by the Council of Europe (1992), North (2007:7) describes that

Level A1 is the point at which the learner can: interact in a simple way, ask and answer simple questions about themselves, where they live, people they know, and things they have, initiate and respond to simple statements in areas of immediate need or on very familiar topics, rather than relying purely on a rehearsed repertoire of (tourist) phrases.

There seems not to be a one to one relation between the descriptors used in the global scale and those for grammatical accuracy.¹³ The discursiveness in the can-do statements at the global level A1 can thus be considered to be quite intermediate and not a beginning learner. Further research needs to be done on the relations between the single subscales.

The other findings by Lenzing and Plesser as well as the present pilot study, underline the issue that CEFR scales are not differentiated enough. It was shown that learners at CEFR level B1 can operate the complex process of feature unification at the sentence level or even the subordinate-clausal level as hypothesized in PT. When abstracting these facts, one can argue that in terms of morpho-syntactic development, the global CEFR scale is not differentiated enough to account for the successive L2 acquisition process as hypothesized in PT. Up to now, discourse-pragmatic features (cf. Topic Hypothesis) and the mapping of causatives, etc. (Di Biase & Kawaguchi 2002; Håkansson, Salameh, & Nettelbladt 2003; Kawaguchi 2005; Kawaguchi, Di Biase, & Pienemann 2005) that are prominent features in the higher CEFR levels cannot be performed by either Rapid Profile or Autoprofiling. More effort, thus, needs to be taken in order to define whether the various subscales in the CEFR account for the detailed developmental path as put forward by PT.

For (b), the difference in developmental stage for learner F04 in AP and RP needs to be discussed. It is worth sharing that this difference might be due to internal features of AP. All subjects got a short introduction into the handling of AP. It was easy to observe that many participants had severe problems with the

13. The Council of Europe argues here that: “this scale should be seen in relation to the scale for general linguistic range shown at the beginning of this section. It is not considered possible to produce a scale for progression in respect of grammatical structure which would be applicable across all languages” (CoE 2001:113). There is however no statement as to the relation between the grammatical accuracy statements and the global scale.

time-constraint embedded in the software. As soon as there is a pause from typing exceeding three seconds, AP deletes the word in the input field. Subject F04 additionally reported that she had little experience with keyboard writing. These relatively long pauses, in which the learner looked for the correct keys to press, triggered the program to delete the still incomplete sentence. This way, the learner was forced to write short sentences/questions that, in terms of syntax, are well under her current state of interlanguage development as assessed with RP. This deletion also caused the data to be much less than the learner would have been able to produce. It also means that there might have been more obligatory contexts for the determination of the developmental stage but that these were deleted due to the narrow time constraint. This fact robbed the learner of the possibility to reach a higher stage. Further, the more advanced learners reported to be quite frustrated with the time they had as they wanted to type in longer, more intricate sentences but AP cut them off in the middle of their production phase. The time constraint is however a very important feature in Auto-Profiling as it rules out any monitoring of the written input (Lin 2012). Reconsidering Plesser's (2008:96) claims about the steadiness of the interlanguage system across mode barrier boundaries, she did not find a time constraint to be extraordinarily crucial: "With regard to time allotment, the learners were asked to finish the tasks within 30 minutes. In fact, the subjects were supposed to be able to solve the tasks within 15–20 minutes, as a result of which they had sufficient opportunity to reread, i.e. monitor their written pieces." As mentioned earlier, her findings suggest that written and oral production rely on the same IL system which has access to both declarative and procedural memory¹⁴ in order to monitor output, IL variation even across mode-barrier boundaries remains within the concept of hypothesis space (cf. Plesser 2008:100ff.). The present study can be assumed to underline the claims about the steadiness of the interlanguage system across mode-barrier boundaries by Plesser (2008) since the results from oral production with RP and written production with AP coincide. Still, further research is required in order to determine whether the time constraint in AP needs to be as narrow as it is administered. More research is needed in order to find out whether AP might be easier to handle for less elaborate, beginning learners who tend to write shorter phrases and sentences.

As for the feasibility aspect in hypothesis (c), it has to be concluded that up to now RP is more feasible than AP. The limitation in feasibility in RP regards the aspect of trained personnel that is needed in order to conduct the profile. Whereas RP needs a researcher to carry out the profile, AP calculates the interlanguage stage fully automatically and online. Lin (2012:54) underlines this by integrating a

14. For further elaboration of procedural and declarative memory, see Ellis (2005).

training section into AP in order to get accustomed to the user interface. AP is still at the beginning of its development and thus needs a certain amount of practice to be feasible for large-scale assessment. Another aspect that is not to be neglected in this regard is the application of the emergence criterion. As mentioned in Section 3.1, in PT, the emergence criterion defines “[...] the beginning of an acquisition process, and focusing on the start of this process will allow the researcher to reveal more about the rest of this process.” (Pienemann 1998: 138). Creative production is thus an essential criterion to pinpoint the acquisition of an underlying interlanguage structure. One must however note that, as mentioned in Section 3.1, for morphology, the structures need to occur in lexical as well as morphological variation in the speech sample (Keßler 2006: 147). The challenge to interlanguage parsers is to take this lexical and morphological variation into account (Lin 2012: 25). Therefore, RP computes distributional analyses and uses implicational scaling. AP follows this approach in “us[ing] the emergence of evidence to determine the learner’s developmental level” (Lin 2012: 30). The problem with AP for research purposes is however that the computation of the structures is hidden. The only reference point for the consistent application of the emergence criterion is the check grammar or all sentences from task feedback. If the few obstacles were, however, to be overcome, AP does have the potential to be integrated into the CEFR to complement it with AP’s advantages as described in the following.

Let us now reconsider as to why a complementation of the CEFR with LP might be useful. Due to the increasing application of CEFR-based ratings as well as the growing importance of valid, reliable and objective means of language testing, this study aimed at complementing some of shortcomings of the CEFR with the computer-based profiling tool that was recently developed. The potential of AP based on PT for this challenge was laid out and the study that pilots this proposition was introduced. Although with a limited field of application, it was shown that AP has the potential to complement CEFR ratings and thus to enhance some of the limitations of the CEFR laid out in part 2.1. The reasons for the potential of AP as a complementation are its rapidness, independence of a present profiler, its high validity and reliability as well as the accuracy and objectivity of the results. The results suggest that AP works about twice as fast as RP does. This advantage lies in the predictive power of Processability Theory in using the emergence criterion as the marking point for the beginning of acquisition. As AP is still being developed and refined, some issues regarding the operability need to be ruled out before it can fully be used in independent testing situations. It was shown that sufficient time for training and familiarization purposes needs to be provided. If this was the case, AP can overcome the obstacles identified in this paper.

It was highlighted that the limited predictive power of the CEFR in terms of beneficial backwash constraints the learner in potential starting points for

self-study. PT-based profiles and instruction can however adhere to this issue in that they give a precise “overview of the learner’s grammatical development at a particular point in time which, moreover, predicts which grammatical structures will be learnable next” (Pienemann & Mackey 1993:135). Furthermore, based on the profile, developmentally moderated feedback (Keßler & Plessner 2011) as well as developmentally guided focus on form instruction (Di Biase 2008) can be provided. With the help of AP, learners are further able to monitor and work on their interlanguage development themselves. The written data used for Autoprofiling is of special interest since up to now, “only some examples of written performance” (CoE 2012) in the CEFR are given. Little (2006) further criticizes that “[...] its illustrative scales are only partly validated by empirical research; the descriptors for written production, for example, were “mainly developed from those for spoken production” (CoE, 2001:220). It can thus be concluded that the written production scale descriptors need to be reviewed.

Due to the study design, there are several limitations that have to be considered. Only a small group of learners was examined. The participants were recruited during their English courses which are divided into the CEFR levels. Some of them were rated by trained CEFR raters, others did the OPT. In this case, the overall or global CEFR levels of the learners were compared to LP. It might have been more suitable to only compare the learners’ grammatical production as this is the only interface that LP is able to complement. However, the learners are not provided with feedback on all the different descriptors which makes it impossible to consult the performance in oral and written production in terms of grammar. Consequently, the global level has to suffice for this purpose. The study group was further not equally represented for a cross-sectional study, the ideal situation is to analyze samples of an equal amount of participants in terms of gender.

As far as the obstacles in the AP analyses are concerned, it is unclear as to whether the very short introduction by the researcher was sufficient enough for the subjects. If the participants had had a chance to familiarize themselves with the program and/or even complete a test run, the difference in the stages of subject F04 might not have occurred.

Even though this study has various limitations, it is hypothesized that a combination of the CEFR and LP is able to broaden the interfaces between LT and SLA (as required by Bachman & Cohen 1998; Shohamy 2000). The paper presented here is a small pilot-study that gives a glimpse into the potentials of a CEFR complementation with PT.

In the framework of PT, the scope-precision dilemma described by Pienemann and Keßler (2007) is worth reexamining. It is hypothesized that using appropriate global tasks, one measurement instrument is able to account for both, scope and

precision. The use of extended communicative information gap activities might be able to provide extensive output not only in terms of grammar but various communicative skills and competencies. This, however, is to be subject to further research.

It would further be helpful to investigate as to where proficiency raters put their focus on while assessing the subjects' levels. Their focus may give a hint to an area of language that is crucial in the rating process. Feasibility studies about the practical integration of AP into the process of rating learners according to the CEFR would shed light on its final operability. The goal of this study was to give insights into areas where both concepts find interfaces that can mutually enhance assessment and language acquisition research.

5. Conclusion

All of the following conclusions have to be considered in relation to the small sample size and are thus to be seen as tentative. My aims in this paper were to discuss whether linguistic profiling can be integrated into ratings based on the CEFR in order to enhance the objectivity, reliability and beneficial backwash¹⁵ of ratings. In order to make inferences about an integration, correspondences between CEFR levels and PT stages had to be determined. The results show that generally, from CEFR level B1, PT stage 5 or higher are assessed. This means that (a) due to the modular approach that PT takes, more correspondence between lower levels of learner language can be found and it was argued that (b) this is due to the accumulated use of discourse features at higher CEFR levels that PT does not capture. It was also hypothesized that (c) CEFR level A1 cannot account for very early phases of language acquisition.

Linguistic profiling within the PT framework has generated two diagnostic tools. RP operates semi-automatically whereas AP elicits the developmental stage online and fully automatically. In order to make claims about the feasibility of each diagnostic tool in rating settings, both were compared in terms of reliability and time allotment. The calculation of the reliability coefficient revealed a high reliability at 94% for the programs. As for the time span that is needed to assess the learners' developmental stages, AP is about twice as fast as RP is. It has to be noted, however, that AP only takes the written learner input into account whereas the oral production in RP in cooperation with a profiler will ultimately take longer.

15. Beneficial backwash refers to the effect of language tests on the teaching and study of the language.

It has thus been concluded that LP has the potential to be integrated into ratings based on the CEFR but that it has a limited scope of application. AP is the diagnostic tool has more potential due to its profiler-independence and rapidness but needs further editing in terms of the user interface and time-constraint as it is still in an early phase of development.

References

- Bachman, L. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Bachman, L. (2004). *Statistical analyses for language assessment* (pp. 153–205, 255–279). Cambridge: Cambridge University Press. doi:10.1017/CBO9780511667350.008
- Bachman L., & Cohen, A. D. (1998). *Interfaces between second language acquisition and language testing research*. Cambridge: Cambridge University Press.
- Bresnan, J. (2001). *Lexical functional syntax*. Malden, MA: Blackwell.
- Brindley, L. F. (1998). Describing language development. In L. F. Bachman & A. D. Cohen (Eds.), *Interfaces between second language acquisition and language testing research* (pp. 112–140). Cambridge: Cambridge University Press.
- Clahsen, H. (1985). Profiling second language acquisition. In K. Hyltenstam & M. Pienemann (Eds.), *Modelling and assessing second language development*. Clevedon: Multilingual Matters.
- Council of Europe. (1992). *Transparency and coherence in language learning in Europe: Objectives, assessment and certification*. Symposium held in Rüşchlikon, 10–16 November 1991. Strasbourg: Council for Cultural Cooperation.
- Council of Europe. (2001). *Common European Framework of Reference for languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press.
- Council of Europe. (2006). *The Common European Framework of Reference in its political and educational dimensions*. Cambridge: Cambridge University Press.
- Council of Europe. (2012). *Illustrations of the European levels of language proficiency*. Accessed on 15 December 2012 from: (http://www.coe.int/t/dg4/linguistic/Manuel1_EN.asp#P81_6526)
- Crystal, D., Fletcher, P., & Garman, M. (1976). *The grammatical analysis of language disability*. London: Arnold.
- Crystal, D. (1982). *Profiling linguistic disability*. San Diego, CA: Singular.
- Di Biase, B., & Kawaguchi, S. (2002). *Exploring the typological plausibility of Processability Theory: language development in Italian second language and Japanese second language*. *Second Language Research*, 18(3), 274–302. doi:10.1191/0267658302slr2040a
- Di Biase, B. (2008). Focus-on-form and development in L2 learning. In J.-U. Kefßler, (Ed.), *Processability approaches to second language acquisition* (pp. 197–220). Newcastle upon Tyne: Cambridge Scholars.
- Ellis, N. (2005). At the interface. Dynamic interactions of explicit and implicit language knowledge. *Studies in Second Language Acquisition*, 27(2), 305–352.
- Ellis, N. (2007). The weak interface, consciousness, and form-focused instruction: Mind the doors. In S. Fotos & H. Nassaji (Eds.), *Form-focused instruction and teacher education. Studies in honor of Rod Ellis* (pp. 197–215). New York, NY: Routledge.

- Håkansson, G., Salameh, E. K., & Nettelblatt, U. (2003). Measuring language development in bilingual children: Swedish- Arabic children with and without language impairment. *Linguistics*, 41, 255–288. doi:10.1515/ling.2003.009
- Håkansson, G., & Norrby, C. (2006). Processability Theory applied to spoken and written L2 Swedish. In F. Mansouri (Ed.), *Second language acquisition research: Theory construction and testing* (pp. 81–94). Newcastle upon Tyne: Cambridge Scholars.
- Harsch, C. (2005). *Der gemeinsame Europäische Referenzrahmen für Sprachen: Leistungen und Grenzen. Bedeutung des Referenzrahmen im Kontext der Beurteilung von Sprachvermögen am Beispiel des semikreativen Schreibens im DESI Projekt*. Unpublished Ph.D. dissertation. University of Augsburg.
- Hymes, D. (1974). *Foundations in sociolinguistics*. London: Tavistock.
- Kaplan, R.M., & Saccuzzo, D.P. (2010). *Psychological testing: Principles, applications, and issues* (8th ed.). Belmont, CA: Wadsworth, Cengage Learning.
- Kawaguchi, S. (2005). Argument structure and syntactic development in Japanese as a second language. In M. Pienemann, *Cross-linguistic aspects of Processability Theory*. Amsterdam: John Benjamins. doi:10.1075/sibil.30.10kaw
- Kawaguchi, S., Di Biase, B., & Pienemann, M. (2005). Extending Processability theory. In M. Pienemann, *Cross-linguistic aspects of Processability Theory*. Amsterdam: John Benjamins. doi:10.1075/sibil.30.09pie
- Kefler, J.-U. (2006). *Englischerwerb im Anfangsunterricht diagnostizieren: Linguistische Profilanalysen am Übergang von der Primarstufe in die Sekundarstufe I*. Tübingen: Narr.
- Kefler, J.-U. (2008). Communicative tasks and second language profiling: Linguistic and pedagogical implications. In J. Eckerth & S. Siekmann (Eds.), *Task-based language learning and teaching. Theoretical, methodological, and pedagogical perspectives* (pp. 291–310). Frankfurt: Peter Lang.
- Kefler, J.-U., & Plessner, A. (2011). *Teaching English grammar* (Standard Wissen Lehramt Englisch). Paderborn: Schöningh/UTB.
- Lado, R. (1961). *Language testing: The construction and use of foreign language tests*. London: Longman.
- Lenzing, A. (2013). *The development of the grammatical system in early second language acquisition. The Multiple Constraints Hypothesis*. Amsterdam: John Benjamins. doi:10.1075/palart.3
- Lenzing A., & Plessner, A. (2010). Challenging the scope-precision dilemma in language testing: The common European framework and linguistic profiling. Paper presented at the 10th International Symposium of Processability Approaches to Language Acquisition (PALA). University of Western Sydney, Australia, 19–21 September.
- Lenzing, A. (2010). *Rapid Profile. A screening procedure for second language acquisition assessment*. Unpublished User Manual.
- Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: The MIT Press.
- Lin, B.J. (2012). *Is automatic linguistic profiling feasible in an ESL context?* Unpublished Ph.D. dissertation. University of Newcastle.
- Little, D. (2008). *The Common European Framework of Reference for languages and the development of policies for the integration of adult migrants*. Council of Europe: Strasbourg. Accessed on 20 February 2013 from: http://www.coe.int/t/dg4/linguistic/Source/Little_CEFRmigrants_EN.doc
- Mackey, A., Pienemann, M., & Thornton, I. (1991). Rapid Profile: A second language screening procedure. *Language and Language Education*, 1(1), 61–82.

- Meisel, J. M., Clahsen, H., & Pienemann, M. (1981). On determining developmental stages in second language acquisition. *Studies in Second Language Acquisition*, 3, 109–135. doi:10.1017/S0272263100004137
- Michell, J. (1999). *Measurement in psychology*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511490040
- North, B. (2007). *The Common European Framework of Reference: Development, theoretical and practical issues*. Accessed on 14 August 2014 from: (<http://www.nationaalcongresengels.nl/cgi-bin/North-Ede-Wagingen%202007-paper.pdf>)
- Neuendorf, K. A. (2002). *The content analysis guidebook*. Thousand Oaks, CA: Sage.
- Pallotti, G. (2007). An operational definition of the emergence criterion. *Applied Linguistics*, 28(3), 361–382. doi:10.1093/applin/amm018
- Pienemann, M., Johnston, M., & Brindley, G. (1988). Constructing an acquisition-based procedure for second language assessment. *Studies in second language acquisition*, 10, 217–243. doi:10.1017/S0272263100007324
- Pienemann, M., & Mackey, A. (1993). An empirical study of children's ESL development and Rapid Profile. In P. McKay (Ed.), *ESL development: Language and literacy in schools, Vol. 2: Documents on bandscale development and language acquisition* (pp. 115–259). Canberra: National Languages & Literacy Institute of Australia and Commonwealth of Australia.
- Pienemann, M. (1998). *Language processing and second language development. Processability Theory*. Amsterdam: John Benjamins. doi:10.1075/sibil.15
- Pienemann, M. (Ed.). (2005). *Cross-linguistic aspects of Processability Theory*. Amsterdam: John Benjamins. doi:10.1075/sibil.30
- Pienemann, M. (2007). Processability Theory. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition: An introduction*. Malden, MA: Blackwell.
- Pienemann, M., & Keßler, J.-U. (2007). Measuring bilingualism. In P. Auer & L. Wei (Eds.), *Handbook of applied linguistics. Vol. 5: Multilingualism* (pp. 247–275). Berlin: Mouton de Gruyter.
- Pienemann, M., Lin, B.-J., & Chung, L.-Y. (2009). The feasibility of auto-profiling online. In E. Damiani et al. (Eds.), *New dimensions in intelligent interactive multimedia systems* (pp. 189–198). Berlin: Springer.
- Pienemann, M., & Keßler, J.-U. (Eds.). (2011). *Studying Processability Theory: Introductory textbook*. Amsterdam: John Benjamins. doi:10.1075/palart.1
- Plesser, A. (2008). *A study on writing and speaking*. Unpublished dissertation. University of Paderborn.
- Plesser, A. (2011). *The interlanguage approach: Example tasks*. Unpublished User Manual.
- Rasinger, S.M. (2008). *Quantitative research in linguistics. An introduction*. London: Continuum.
- Rohrmann, B. (2007). *Verbal qualifiers for rating scales: Sociolinguistic considerations and psychometric data*. University of Melbourne, Australia. Accessed on 14 August 2014 from: (<http://www.rohrmannresearch.net/pdfs/rohrmann-vqs-report.pdf>)
- Schaefer, E. (2008). Rater bias patterns in an EFL writing assessment. *Language Testing*, 25, 465–493. doi:10.1177/0265532208094273
- Shohamy, E. (2000). The relationship between language testing and second language acquisition, revisited. *System*, 28, 541–553. doi:10.1016/S0346-251X(00)00037-3
- Stevens, S. S. (1946). On the theory of scales of measurement. *Science*, 103, 677–680. doi:10.1126/science.103.2684.677
- Wigglesworth, G. (1993). Exploring bias analysis as a tool for improving rater consistency in assessing oral interaction. *Language Testing* 10, 305–319. doi:10.1177/026553229301000306

- Wilkins, D. A. (1978). Proposal for levels definition. In J. L. M. Trim (Ed.), *Some possible lines of development of an overall structure for a European unit/credit scheme for foreign language learning by adults* (pp. 71–78). Strasbourg: Council of Europe.
- Winke, P., Gass, S., & Myford, C. (2013). Raters' L2 background as a potential source of bias in rating oral performance. *Language Testing*, 30, 231–252. doi:10.1177/0265532212456968

Appendix

Results Oxford English Placement Test ©	Recommended Course
0	English I+II (A1)
1–19 (A1)	English III+IV (A2)
20–39 (A2)	English V+VI (B1)
40–49 (B1.1)	English VII+VIII (B1/B2)
50–59 (B1.2)	English IX+X (B2)
60–69 (B2.1)	English XI+XII (B2/C1)
70–79 (B2.2)	English XIII+XIV (C1)
80–89 (C1.1)	English XV+XVI (C1/C2)
> 89 (C1.2)	English XVII+XVIII (C2)
from 60 points (B2)	Listening & Speaking
from 60 points (B2)	Reading & Writing

Figure 11. OPT and recommended English course; available from: <http://kw.uni-paderborn.de/institute-einrichtungen/zfs/sprachkurse/englisch/>

Assessing linguistic levels of L2 English in primary school programs

Esther Maier, Lea Neubauer, Katharina Ponto,
Stefanie Couve de Murville & Kristin Kersten
University of Hildesheim

This paper compares the outcomes of different studies on the L2 acquisition of English in different primary school settings within the framework of Processability Theory (PT). The results show that children from immersion (IM) programs tend to reach higher stages compared to pupils from traditional teaching programs. The intensity and the duration of L2 classroom contact show the strongest effect on the test results. Prior experience with the L2 before primary school also has a significant effect, while sex, age and home language use do not influence the level of attainment reached by the learners. In addition, the study investigates the suitability of linguistic profiling for highly advanced young learners of English, as well as the communicative tasks used for data elicitation. Recommendations for an adaptation of the tasks are derived from the observations.

1. Introduction

The acquisition of a second language in addition to one's mother tongue/s is a phenomenon which has become increasingly necessary in the globalized world of today (Larsen-Freeman & Long 1991:2). This is also recognized within the European Commission's promotion of the *mother tongue + 2 languages*-principle (EU Commission 2003), which calls all European citizens to strive for bilingual or multilingual competence during their formative years. Within the 23 officially recognized languages in the European Union, English has a special position as Europe's lingua franca as the "most taught foreign language in nearly all countries at all educational levels" (Eurydice 2012:11). The acquisition of a second language occurs due to and in combination with various factors, in differing circumstances, and mediated by different teaching strategies. The identification of beneficial circumstances and teaching strategies and, with this, the assessment of English language competence, have gained special significance in school

curricula, starting at a very young age, i.e. in preschool and primary school (EU Commission 2003).

However, the types of teaching programs and the language competences they foster are very heterogeneous across Europe. This is also true for Germany, where English is the first foreign language learners are introduced to in almost all federal states (KMK 2005:6ff).

As research has shown, consideration should be given to a number of factors to account for the effectiveness of such teaching programs. Among these are, for instance, external factors such as the age of onset (Johnstone 2002; Long 2007; Muñoz 2011), the duration and intensity of L2 input (Wesche 2002; Weitz et al. 2010), the language use (L1 and L2) at home and at school (e.g. Piske et al. 2001; Housen et al. 2011), and the quality of L2 input (Snow 1990; Massler, & Ioannou-Georgiou 2010; Weitz et al. 2010), as well as social factors (Ellis 2008) and internal factors (Lightbown & Spada 1999). However, all these factors form a very complex grid of interrelated and interacting variables (Ellis 1985), a fact which needs to be considered in second language acquisition research. To date, on the basis of current research no comprehensive overview is available of the level of competence that can be reached at the end of programs which differ with regard to such a variety of factors.

The purpose of this study is to compare the linguistic levels that learners of L2 English reach in different EFL (English as a Foreign Language)¹ primary programs in Germany, ranging from traditional foreign language teaching with two hours per week starting in grade 3, to partial IM programs where 50% or more of the curriculum is taught in the foreign language (L2) starting in grade 1.² In addition, the study investigates the influence of some of the above-mentioned factors (i.e. duration and intensity of L2 classroom input, prior experience with the L2, age, sex and home language background) on the participants' language levels. Being aware of the complexity of influencing factors (Ellis 1985), only a selection of relevant factors can be taken into account within the context of this study. The linguistic data were elicited through communicative tasks (Long 1985:94; Keßler & Kohli 2006), assessed with the help of *Rapid Profile* (Keßler & Liebner 2011), and discussed with respect to the framework that each of the schools offers.

1. The terms *second language* and *foreign language* are used interchangeably in this article, and are both subsumed under the abbreviation *L2*.

2. The data and parts of the text of this study are based on studies carried out by Maier (2011), Neubauer (2013), Schwirz (2012), Wenzel (2011), and Wiegand (2012).

The suitability of the data elicitation and assessment procedures used in the study³ is also discussed, especially with regard to the high language competence encountered in the IM classes. It is hoped that the results will present a basis of comparison for future studies into the language levels that can be expected in different school programs, and, ultimately, into the effect of different factors which are claimed to be beneficial for L2 development. However, to arrive at a generalizable conclusion, a much wider database is needed.

2. Child data studies within the PT framework

The hierarchy of the morpho-syntactic development of EFL has been the focus of numerous studies. Studies by Pienemann and Mackey (1993), Pienemann, Keßler, and Liebner (2006) and Kersten (2009a) are referred to, the findings of which are compared to research recently conducted by Maier (2011), Neubauer (2013), Schwirz (2012), Wenzel (2011), and Wiegand (2012).⁴

In his pioneering study on the language development of adult learners of EFL, Johnston (1985) confirmed that the learners follow certain developmental sequences predicted by PT and that the hypothesized sequences are implicational. The study by Pienemann and Mackey (1993) was the first to support the predictions posited by PT with respect to the age group of young learners. This study was conducted with 13 children aged eight to ten years (Pienemann 1998:179). As in the Johnston study (1985), the results of Pienemann and Mackey (1993) support the hypothesized implicational pattern, although the data base was not as rich and, thus, the internal consistency was less strong than in the previous study by Johnston (1985). Nevertheless, both studies provide evidence in support of the processability hierarchy and confirm the implicational pattern inherent in the theory (Pienemann 1998:180) for different age groups.

Other more recent studies support these findings for children of a similar age group. Within the context of the MILES-project, Pienemann et al. (2006:67)

3. We are very grateful to Susanne Kurth for providing the pictures for the communicative tasks, to Lisa Bade, Nora Chihabi, Annika Ellermann, Nicole Gnewuch, Stefanie Hartmann, Anna Rebekka Hintz, Aylissa Hoffmann, Martin Preisigke, Annika Schmidt and Jana Wiegand for help with data collection and analysis, and to Dario Klemm for the statistical analysis, as well as to all teachers and children who have been involved in the project. Without their help, this study would not have been possible.

4. The studies were conducted within the research project *Multilingualism in Preschool and Primary School* at Hildesheim University, headed by Kristin Kersten.

analyzed speech samples from 70 pupils learning English in different settings. In order to investigate whether there are general observable patterns of how English is learnt and, if so, what levels of attainment the children are actually able to reach, four different learner groups were tested with the help of several communicative language tasks (Pienemann et al. 2006:68). These groups consisted of 16 learners from a mainstream German school learning their L2 via the so-called *Begegnungskonzept*,⁵ 28 children learning English in Sweden, 12 learners from a partial IM primary program in Germany (six – grade 1, six – grade 3) and 14 learners from secondary school. In a second step, the oral data were analyzed and examined according to the linguistic structures predicted by PT (ibid.). The results of this analysis showed that the developmental stages proposed by PT could be confirmed for all of the primary school learners. Tables 1 and 2 show the results of the *Begegnungskonzept* and the IM groups, which are relevant for comparison with the data of this study (cf. Section 3).⁶

Table 1. Results of Grade 4 – *Begegnungskonzept* in Mainstream Primary School (approx. 8% L2 Intensity) (adapted from Pienemann, Keßler, & Liebner 2006: 86) (+: emerged)

Participants		C	C	C	C	C	C	C	C	C	C	C	C	C	C	C
L2 contact (months)		24	24	24	24	24	24	24	24	24	24	24	24	24	24	24
Begegnungskonzept	6															
	5															
	4															
	3															+
	2	+	+						+		+	+	+			
	1	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+

5. This concept was implemented in North Rhine-Westphalia in 1992 and was the first step of an early beginning of learning a second language (Beckmann 2006:19). The focus of this concept is on conveying intercultural competence and a positive attitude towards the L2 rather than on teaching specific language rules (Leopold-Mudrack 1998:15).

6. For more detailed information of the study and the results of the other groups, see Pienemann, Keßler, & Liebner (2006).

Table 2. Results of Grades 1 and 3 in an Immersion Program (approx. 62% L2 Intensity) (adapted from Pienemann, Keßler & Liebner 2006: 86) (+: emerged)

Participants		IM 40	IM 41	IM 42	IM 43	IM 44	IM 45	IM 74	IM 75	IM 76	IM 77	IM 78	IM 79
L2 contact (months)		11	11	11	11	11	11	33	33	33	33	33	33
Immersion Groups	6												
	5							+	+				
	4							+	+	+	+	+	+
	3	+	+				+	+	+	+	+	+	+
	2	+	+				+	+	+	+	+	+	+
	1	+	+	+	+	+	+	+	+	+	+	+	+

In 2009, Kersten collected speech samples of four children attending a partial IM primary school in Kiel, Germany (Kersten 2009a:268). The longitudinal study elicited data at the end of each school year from grade 1 to grade 4 (Kersten 2009a:269). With the help of a picture story, the speech samples were elicited and later analyzed based on the operational criteria for data coding adapted from Pienemann (1998) and Pallotti (2003, 2007) (Kersten 2009a:283). Table 3 shows the different attainment levels of each of the four participants at the end of each grade.

Table 3. Results of Grades 1–4 in an Immersion School (approx. 70% L2 Intensity) (adapted from Kersten 2009a: 289) (+: emerged; (+): insufficient evidence for emergence; (-): insufficient evidence against emergence; -: not emerged; /: no evidence)

Participants		3.1	3.2	3.3	3.4	6.1	6.2	6.3	6.4	7.1	7.2	7.3	7.4	8.1	8.2	8.3	8.4
L2 contact (months)		10	22	34	46	10	22	34	46	10	22	34	46	10	22	34	46
Immersion	6									+		(+)	(+)		(+)	+	+
	5		+	+	+	(+)	+	+	+		+	+	+	-	+	+	+
	4		(+)	(+)	(+)	(+)	+	+	+	(-)	+	+	+		(+)	+	+
	3	(+)	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
	2	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
	1	/	/	/	/	/	/	/	/	/	+	/	/	/	/	/	/

According to Kersten (2009a:291), “all stages predicted by PT could be confirmed” within this study. The comparison to the group of naturalistic L2 learners

(Pienemann & Mackey 1993) revealed that the participants reached comparable levels of attainment in L2 acquisition (*ibid.*).

However, it should be noted that the comparability of the grammatical language competence in different programs is highly dependent on the setup and the characteristics of the respective programs. Indeed, factors like intensity and duration of the L2 exposure, the emphasis on content and language, and the implementation of EFL teaching principles may vary between programs. As IM teaching, an intensive form of CLIL (Content and Language Integrated Learning) (Mehisto 2011: 121), makes use of the L2 as the medium of instruction, the intensity of L2 exposure is much higher in IM programs than in traditional EFL programs. With regard to the IM programs referred to by Pienemann et al. (2006: 69) and by Kersten (2009a: 268), the amount of L2 exposure is approximately 70% of obligatory teaching hours, while mainstream schools mostly include two English lessons per week providing a much lower intensity of L2 exposure (Piske 2006: 1).

3. The study

3.1 Research questions

This study focuses on L2 learners from different IM programs and from a mainstream EFL class. Applying PT to these different settings, four research questions are relevant for the investigation:

1. What stages do the participants reach in the different school programs?
2. What factors influence these results?
3. Are the stages predicted by PT suitable to assess learners in primary IM programs?
4. How suitable are the communicative tasks for assessing the L2-level of pupils in different teaching programs?

Based on the evidence suggested by earlier studies, *i.e.* Pienemann et al. (2006) and Kersten (2009a), the following hypotheses are proposed:

- (H1) It is expected that IM pupils reach stages 4, 5 and 6 at an early grade. Learners from a traditional approach are expected to reach stages 2 and 3.
- (H2) It is expected that the duration of the L2 contact and its intensity are the most important predictors of the stages reached in the different programs.
- (H3) If IM pupils already reach high stages from the first year on, the PT framework might not be able to differentiate the high proficiency competences of intensive program learners.

- (H4.1) The communicative tasks will not be equally suitable for all levels of proficiency.
- (H4.2) It will be difficult to provide sufficient obligatory contexts within the tasks for all stages relevant to the PT hierarchy.

3.2 Schools

3.2.1 *School A*

School A is an early partial IM school with approximately 160 pupils. The private school consists of a kindergarten, a primary school and a secondary school. The primary school has two parallel classes for the grades 1–3 and one class for grade 4. A secondary level was implemented three years ago and comprises one class for grades 5, 6 and 7. The concept of the school derives from the Canadian IM programs. The school teaches according to the core curriculum of Lower Saxony (*Niedersächsisches Kultusministerium 2006*) and, in addition, to the Cambridge International Curriculum, combining both to form its own curriculum. All subjects except for German are taught in English, which amounts to an average of 67% of teaching hours. All of the teachers are either native speakers of English or highly competent speakers of English.

3.2.2 *School B*

School B is a state primary school with approximately 200 pupils and has a strong emphasis on English and Physical Education. Within each year group, one of three classes is an IM class. The early partial IM program has been applied since 2008. As in school A, all subjects except for German are taught in English. The intensity of the IM program ranges from 69–73% of obligatory teaching hours according to grade. For pupils attending before and after school day care the average is even higher.

All of the IM teachers studied English as part of their teaching qualifications. One of the IM teachers is a native speaker of English, three of the IM teachers are highly competent speakers of English. Throughout the school day, the entire staff maintain the “one-person one-language principle” (Döpke 1992).

3.2.3 *School C*

School C is a mainstream primary school of approximately 145 pupils, with two classes per year group. Since August 2008, the school has been a full-time school and also offers afternoon childcare. One of the characteristics of the school is the multicultural background of the pupils; the diversity of the school is considered an enrichment to the learning environment.

Unlike school A or B, the school does not follow an IM program. Instead, English is taught for two hours per week, as is common in most mainstream primary schools in Germany. The English teachers are native speakers of German. The teacher of the participating class is not a trained teacher of English.

3.3 Participants

In the study, 105 primary school participants were tested according to the hierarchy described by PT. Table 4 represents the summary of the most relevant information of the participants' background assumed to influence the stages reached by the learners. This information was collected via a questionnaire handed out to the participants' parents. The tests included three groups from school A ($n=47$, A1=grade 1, A2/A3=grade 4), two groups from school B ($n=44$) which were tested twice (B1=grade 2 (t_1) & 3 (t_2), B2=grade 3(t_1) & 4(t_2)) and one group from school C (C1=grade 4, $n=14$). The sex ratio among the groups is balanced as 55 female and 50 male participants were tested. The age differs as the groups belong to different grades. The L2 exposure of group C1 is lower than the other participants' exposure because the school starts their foreign language classes at the beginning of grade 3. Therefore, the L2 classroom exposure in months of Groups A1 and C1 are comparable. Group A2 started with the L2 at the beginning of the second school year. The other groups (A3, B1, B2) started with EFL at the beginning of primary school. Therefore, the average L2 exposure varies from 10/13 months (A1, C1), to 45 months (B2 at t_2).

Approximately half of all participants (48) did not have prior experience in the L2, while a total of 38 learners had intensive prior experience, such as staying abroad longer than nine months or visiting a bilingual preschool. The latter applies to 26 of the 38 participants with prior experience. For 77 participants, German is the home language, while 19 participants use other languages besides German.⁷

3.4 Method

3.4.1 Data elicitation

Data elicitation was carried out individually in a separate room of the school building. All tests were video- or audiotaped. Two interviewers, one speaker of German (IG), the other of English (IE), carried out the tests. After an introduction, four communicative tasks were used for eliciting the individual speech samples. The

7. As not all questionnaires were returned, this information could not be reported for all participants.

Table 4. Overview of Participants (n.i. – no information)

Group	Grade	Number of Participants	Sex		Average Age (months)/ Age Range			Average L2 Classroom Exposure (months)	L2 Intensity at School	Prior L2 Experience				Language Use			
			f	m	t ₁	t ₂	No			Non-intensive	Intensive	n.i.	German	Multi-lingual	n.i.		
A1	1	11	5	6	85	74–92	–	10	67%	3	–	2	6	4	2	5	
A2	4	14	10	4	120	112–127	–	30,3	67%	11	2	1	–	14	–	–	
A3	4	22	6	16	121	109–132	–	37,9	67%	9	4	2	7	17	5	–	
B1	2/3	23	11	12	97	91–108	109	103–120	22/34	69/72%	8	1	12	2	15	6	2
B2	3/4	21	13	8	109	101–131	121	113–143	33/45	72/73%	7	4	9	1	18	2	1
C1	4	14	10	4	116	106–122	–	14	8%	10	1	–	3	9	4	1	

design, as well as the structures which were intended to be elicited within each task, are presented in the following section.

3.4.2 Communicative tasks

Communicative tasks may “provide a natural context for language production” (Keßler & Liebner 2011: 141) and are, therefore, well-suited for the screening procedure. Keßler and Liebner (2011: 142) claim that it is essential to choose tasks “which involve the informants producing those structures that are relevant for the screening”. Put simply, when the use of certain linguistic structures is intended, it has to be made sure that the tasks used for data elicitation offer enough contexts for these very structures.

3.4.2.1 Structured Interview. The first task was a structured interview. Through this task, the participant got to know the interviewers better and an atmosphere of trust was created to alleviate anxiety associated with testing. First, the participant was interviewed by the IG in German about general topics such as friends and hobbies. Then s/he was prompted to switch into the role of the interviewer and had to ask the English-speaking interviewer (IE) the same questions. Obviously, the participant had to use English for interviewing the IE, which is why this task is suitable for eliciting a number of question structures (e.g. “Where do you live?”, “Do you have any pets at home?”).

3.4.2.2 Picture Difference Task. The structured interview was followed by a picture difference-task: the learner received a picture which corresponded to the picture of the interviewer but lacked a number of elements, such as a *sun*, a *big girl* or *two birds*, etc. (Figure 1). These elements were given to the participant

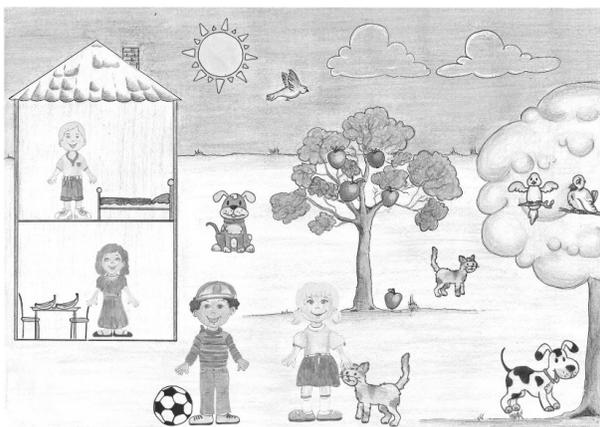


Figure 1. Picture Difference-Task (Complete Version of IE; Original Picture in Color)

separately. By doing so, an information gap was created (Keßler & Kohli 2006: 93), which increases the communicative value of the task. In order to find the correct place of the elements in the picture of IE, the learner had to use a number of specific question patterns, like copula verb-inversions (“Is the apple on the table?”) or Wh-copula-questions (“Where are the bananas?”).

3.4.2.3 Storytelling/Story Completion Task. The next task consisted of a picture story about a platypus that escapes from the zoo. A boy, on finding the platypus, wonders where the animal might have come from, and finally returns it to the zoo. The participant received the second part of the story, was given time to grasp it, and was then supposed to tell the the story to the IE with the help of the pictures. Afterwards, the learner had the opportunity to ask the IE questions about the beginning of the story.⁸

This task primarily aimed at eliciting SVO word order, past tense (IG gave the information that the story took place “last week”) and free questions. In contrast to the other three tasks, this task also provides contexts for the formation of cancel inversion in pictures which show that the boy asks himself questions about the habits of the platypus, indicated by several thought bubbles (Figure 2).

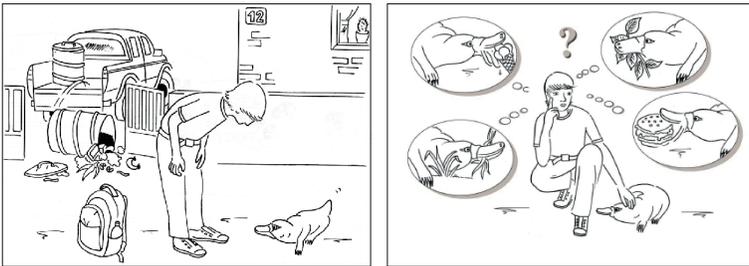


Figure 2. Platypus Picture Story (Pictures 5 and 6)

3.4.2.4 Habitual Action Task. The final part of the test was a habitual action task. Again, a number of pictures were given to the learner. This time, however, a typical day in the life of a young girl was illustrated. It was the participant’s task to describe the daily routines of this girl. Therefore, the IG instructed the learner to begin with “every day”, so that a context for the 3rd person singular -s was given. Apart from that, SVO was elicited. An example of pictures used for this task is given in Figure 3.

8. This was only possible at the first time of elicitation. A year later, the story completion part had to be left out as the children were already familiar with the beginning of the story.

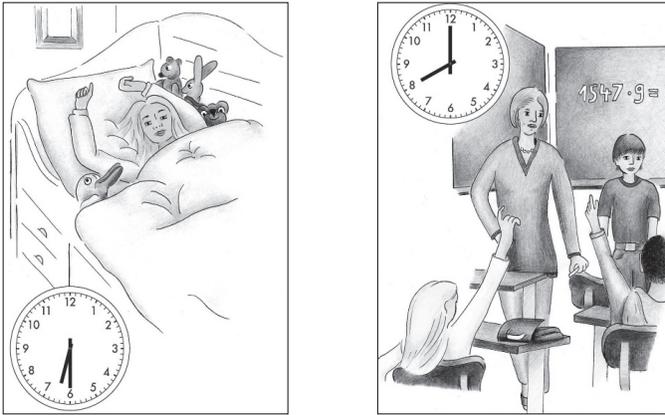


Figure 3. Habitual Action-Task (Pictures 1 and 4, Original Pictures in Color)

To determine whether the tasks were suitable for the different groups of participants and if adaptations were needed, it was necessary to identify criteria for their suitability. The criteria used for this analysis were:

1. whether the tasks “provide a natural context for as many relevant structures possible in a short time” (Keßler & Liebner 2011:142): in other words, the test has to contain tasks that elicit a sufficient amount of linguistic structures which have to be used in order to solve them;
2. whether the task is motivating: according to Kyriacou (1997:26), “motivation involves an interest in the learning task itself and also satisfaction being gained from the task”. With regard to the current study, this is indicated by an active engagement with the task and implies sustained attention to the task. Rejection, long pauses and obvious disinterest, on the other hand, would be clear indicators for the tasks being non-motivating;
3. whether the task is easily comprehensible for the children: the degree of autonomy with which a task is carried out serves as an indicator for comprehension. An easily comprehensible task enables the learner to work with it without the need for further explanations.

3.4.3 Data analysis

Following data collection, the speech samples were transcribed and analyzed according to the PT stages. Direct repetitions or structures which were influenced by the interviewer, acoustically unclear morphological endings, uninflected verb forms and non-target like structures that the hierarchy does not account for

(e.g. “Where comes...?”, “Read you a book?”) were excluded from the analysis. Apart from that, passive structures were also neglected, as these structures are not (yet) implemented in the developmental schedule that PT proposes. Afterwards, the structures were entered into Rapid Profile (a computer-assisted screening procedure developed to analyze speech data, which underlies the principles of PT, cf. Keßler & Liebner 2011: 133ff).

The results were then analyzed statistically, with special regard to the effect size of different variables which are claimed to have an impact on the L2 competence of the participants, i.e. sex, age, prior experience with the L2, home language use, L2 contact duration, and L2 intensity (cf. Introduction).

4. Results

4.1 Stages

4.1.1 School A

In the first group in School A (A1), two of the participants reached stage 3 of the developmental hierarchy, four reached stage 4, another four participants reached stage 5 and one participant reached stage 6 (Table 5). The median of group A1 is stage 4 (Figure 4).

Table 5. Results of Group A1 (Grade 1, IM approx. 67% L2 Intensity) (Wenzel 2011) (+: emerged; (+): insufficient evidence for emergence; /: no evidence)

Participants		A 15	A 16	A 17	A 18	A 19	A 20	A 21	A 22	A 23	A 24	A 25
L2 contact (months)		10	10	10	10	10	10	4	10	10	10	10
Group A1 (Grade 1) Stages	6	+										
	5	(+)		+	+	+				+		
	4	+	+	+	+	+		+	+	+		+
	3	+	+	+	+	+	+	+	+	+	+	+
	2	+	+	+	+	+	+	+	+	+	+	+
	1	+	+	+	+	+	+	+	+	+	+	+

In the second group of School A (A2), three of the participants reached stage 4, ten stage 5 and one participant stage 6 (Table 6). The median of group A2 is stage 5 (Figure 4).

4.1.2 School B

The participants of School B were tested twice with an interval of a year between the tests (Table 8 – for group B1, Table 9 – for group B2). In the first testing (t_1) of group B1, 17 of the participants reached stage 5 and four participants reached stage 6. In the second testing (t_2), five participants reached stage 5 and twelve participants reached stage 6. The median of the first testing is stage 5 and stage 6 in the second testing (Figure 4).

Table 8. Results of Group B1 at Test Times 1 & 2 (Grade 2 (t_1) & 3 (t_2), approx. 69/72% L2 Intensity) (+: emerged (+): insufficient evidence for emergence; /: no evidence)

Group B1		Grade 2 – t_1						Grade 3 – t_2							
Participants		L2 contact (months)	Stage						L2 contact (months)	Stage					
			1	2	3	4	5	6		1	2	3	4	5	6
B	1	22	+	+	+	+	+	+	–						
B	2	22	+	+	+	+	+		34	+	+	+	+	+	
B	3	22	+	+	+	+	+	+	34	+	+	+	+	+	+
B	4	22	+	+	+	+	+		34	+	+	+	+	+	+
B	5	22	+	+	+	+	+	+	34	+	+	+	+	+	+
B	6	22	+	+	+	+	+		34	+	+	+	+	+	+
B	7	22	+	+	+	+	+		34	+	+	+	+	+	+
B	8	22	+	+	+	+	+		34	+	+	+	+	+	+
B	9	22	+	+	+	+	+		34	+	+	+	+	+	+
B	10	22	+	+	+	+	+		–						
B	11	22	+	+	+	+	+		34	+	+	+	+	+	
B	12	–							34	+	+	+	+	+	+
B	13	22	+	+	+	+	+		34	+	+	+	+	+	+
B	14	22	+	+	+	+	+		34	+	+	+	+	+	
B	15	–							34	+	+	+	+	+	+
B	16	22	+	+	+	+	+		–						
B	17	22	+	+	+	+	+	+	–						
B	18	22	+	+	+	+	+		34	+	+	+	+	+	+
B	19	22	+	+	+	+	+		34	+	+	+	+	+	
B	20	22	+	+	+	+	+		34	+	+	+	+	+	+
B	21	22	+	+	+	+	+		34	+	+	+	+	+	
B	22	22	+	+	+	+	+		–						
B	23	22	+	+	+	+	+		–						

In the first testing (t_1) of group B2, two participants reached stage 4, thirteen reached stage 5 and three reached stage 6. In the second testing (t_2), four participants reached stage 4, five reached stage 5 and twelve reached stage 6.⁹ The median of the first testing is stage 5, and stage 6 in the second testing (Figure 4).

Table 9. Results of Group B2 at Test Times 1 & 2 (Grade 3 (t_1) & 4 (t_2), approx. 72/73% L2 Intensity) (+: emerged; (+): insufficient evidence for emergence; /: no evidence)

Group B2		Grade 3 – t_1						Grade 4 – t_2							
Participants		L2 contact (months)	Stage						L2 contact (months)	Stage					
			1	2	3	4	5	6		1	2	3	4	5	6
B	24	33	+	+	+	+	+		45	+	+	+	+	+	+
B	25	33	+	+	+	+	+		45	+	+	+	+	+	+
B	26	–							45	+	+	+	+	+	+
B	27	33	+	+	+	+	+	+	45	+	+	+	+	+	+
B	28	33	+	+	+	+	+		45	+	+	+	+	+	+
B	29	33	+	+	+	+	+		45	+	+	+	+	+	+
B	30	33	+	+	+	+	+		45	+	+	+	+		
B	31	33	+	+	+	+	+		45	+	+	+	+	+	+
B	32	33	+	+	+	+	+		45	+	+	+	(+)	+	+
B	33	33	+	+	+	+	+	+	45	+	+	+	+	+	+
B	34	33	+	+	+	+	+		45	+	+	+	+	+	
B	35	33	+	+	+	+			45	+	+	+	+	+	
B	36	33	+	+	+	+	+		45	+	+	+	+	+	+
B	37	33	+	+	+	+			45	+	+	+	+		
B	38	–							45	+	+	+	+	+	
B	39	33	+	+	+	+	+		45	+	+	+	+		
B	40	–							45	+	+	+	+		
B	41	33	+	+	+	+	+		45	+	+	+	+	+	
B	42	33	+	+	+	+	+	+	45	+	+	+	+	+	+
B	43	33	+	+	+	+	+		45	+	+	+	+	+	
B	44	33	+	+	+	+	+		45	+	+	+	+	+	+

9. For more information on the second set of data of group B2 see Wiegand (2012).

4.1.3 School C

In the group of participants from School C (C1), which teaches English following a traditional EFL approach, nine of the participants reached stage 2, two reached stage 3 and three stage 4 (Table 10). The median of group C1 is 2 (Figure 4).

Table 10. Results of Group C1 (Grade 4 Traditional Program approx. 8% L2 Intensity) (+: emerged; (+): insufficient evidence for emergence; /: no evidence)

Participants		C 1	C 2	C 3	C 4	C 5	C 6	C 7	C 8	C 9	C 10	C 11	C 12	C 13	C 14	
L2 contact (months)		13	13	13	13	13	13	13	13	13	13	13	13	13	13	
Group C1 (Grade 4)	Stages	6														
		5														
		4			+			+	+							
		3			/			+	(+)			+		+		
		2	+	+	+	+	+	+	+	+	+	+	+	+	+	+
		1	+	+	+	+	+	+	+	+	+	+	+	+	+	+

Figure 4 shows the averages of the stages reached by participants in each of the six different groups, compared with the results from earlier studies on child language development (cf. Section 2, Pienemann et al. 2006; Tables 1 and 2, Kersten 2009, Table 3) (H1).

4.2 Comparison among grades

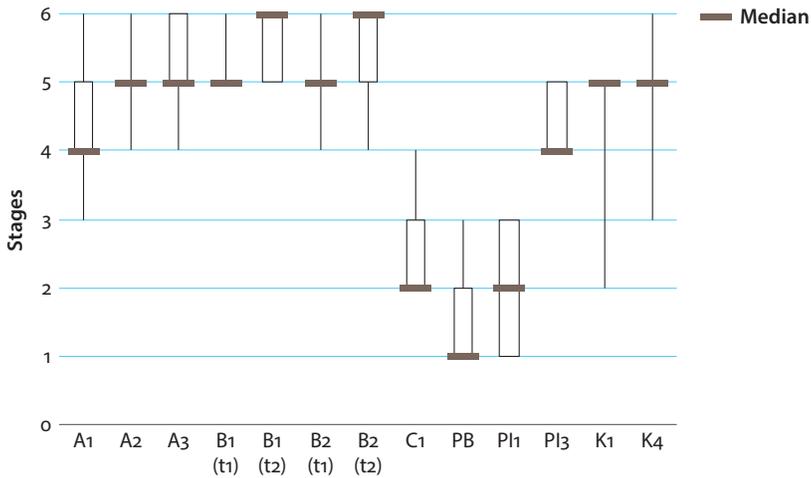
4.2.1 Comparison of Grade 1 Groups

Comparing the two IM grade 1 classes (A1 and Pienemann et al. 2006), an ANOVA reveals that there is a significant difference between the A1 group and the group from Pienemann et al. (2006) (Figure 4).¹⁰

4.2.2 Comparison of Grade 3 Groups

With regard to the results in all IM grade 3 classes (B1 (t_2), B2 (t_1) and Pienemann et al. 2006; Figure 4), a statistical comparison shows significant differences between B1 (t_2) and B2 (t_1) as well as between B1 (t_2) and the IM group from grade 3. A comparison between B2 (t_1) and the IM group from grade 3 shows no significant differences.

10. An ANOVA showed significant differences between the groups ($F(5.91) = 93.216$, $p = .00$).



Group	A1	A2	A3	B1 (t1)	B1 (t2)	B2 (t1)	B2 (t2)	C1	PB	P1	P3	K1	K4
Grade	1	4	4	2	3	3	4	4	4	1	3	1	4
N	11	14	22	21	17	18	21	14	16	6	6	4	4
Programme (Intensity)	67%	67%	67%	69%	72%	72%	73%	8%	8%	62%	62%	70%	70%
Duration (Months)	10	30,3	37,9	22	34	33	45	13	24	11	35	10	46
Mode	4,5	5	5	5	6	5	6	2	1	3	4	5	5

Figure 4. Stages reached across groups from this Study (A1-3, B1-2, C1), Pienemann, Keßler & Liebner (2006, PB *Begegnungskonzept* group, P1 immersion group grade 1, P3 immersion group grade 3), and Kersten (2009a, K1 immersion group grade 1, K4 immersion group grade 4)

4.2.3 Comparison of Grade 4 Groups

Statistical comparison of all grade 4 classes (A2, A3, B2, C1, Kersten 2009a, Pienemann et al. 2006) reveals that there are significant differences between the traditional programs (C1 and the *Begegnungskonzept*) on the one hand, and the IM programs (A2, A3, B2 and Kersten 2009a) on the other hand. No significant differences were found among the grades 4 classes of the IM programs (A2, A3, B2, Kersten 2009a). In addition, the difference between the two traditional programs, C1 and the *Begegnungskonzept* (Pienemann et al. 2006), was also significant.

4.2.4 Longitudinal Development

B1 and B2 were tested twice with an interval of a year between the tests (B1 in grades 2 and 3, B2 in grades 3 and 4). A repeated measure analysis shows a significant increase in the development from the first to the second testing for both groups.¹¹ The strength of the increase from t_1 to t_2 did not differ significantly between B1 and B2.

11. Repeated measure analysis for B1: $F(1.15) = 16.0, p = .001$; for B2: $F(1.18) = 5,591, p < .05$.

4.3 Factors affecting L2 development

To determine the influence of different factors which are hypothesized to have an effect on the results (cf. Introduction), the data of the heterogeneous groups of this study were pooled (with results from B1 and B2 at t_2), clustered and analyzed statistically. Due to the rather small size, the sample is not considered representative. A parents' questionnaire was used to gather background information (cf. Section 3.3). On the basis of the categories and questions elicited in the parent's questionnaire, information was available for the following factors: sex, age, L2 contact duration, L2 intensity, prior L2 experience, and home language use. The latter factor indicates whether a child exclusively uses German at home in conversation with their family, or another language as well (cf. also Section 3.3).¹² These factors were considered with regard to their effect sizes for the stages reached.

The data of 80 participants included information on all relevant variables and was used for the analysis. The cluster analysis suggested three clusters. A MANOVA revealed that differences were significant for L2 contact and L2 intensity ($p = .00$), and significant for prior experience ($p < .05$) (cf. H2). No differences were found for the factors sex, age and home language use. Figure 5 illustrates the different effect sizes for these variables.

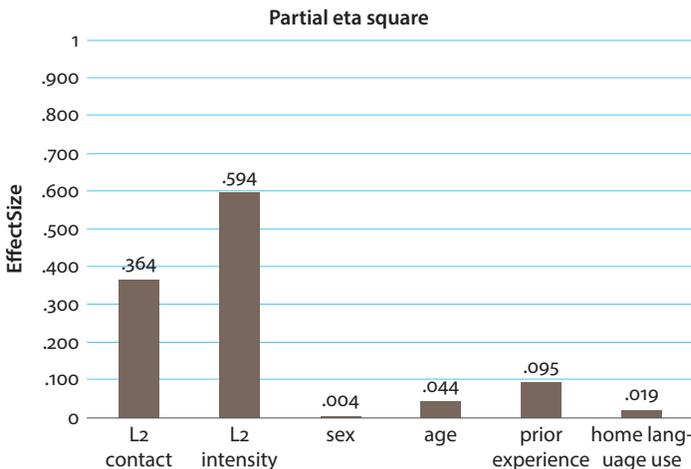


Figure 5. Effect Size of Different Factors Affecting the Test Results from Schools A, B (t_2), C ($n=80$)

12. The factor *home language use* was introduced in order to avoid the category “migration background“, which does not necessarily include information relevant to the specific language use of the child in interaction with family members.

4.4 Communicative tasks

4.4.1 *Sufficient elicitation of structures (cancel inversion)*

Within both the IM contexts and the traditional primary school contexts, the data showed that the tasks elicit sufficient relevant linguistic structures. With regard to cancel inversion, however, some of the pupils did not make use of the four contexts in both pictures, but rather described the four possibilities with one sentence.

(1) A 08: then he thought where he will live

As a consequence, in many interviews only two cancel inversion structures were elicited. Given the fact that the emergence criterion requires three instances of a structure for syntax (Pienemann et al. 2006: 78), there was one context missing and the highest stage was thus not counted as acquired for a number of participants.

Due to this observation, Schwirz (2012) suggested the modification of the picture story by adding a third picture which provides an additional context for cancel inversion (in which the boy wonders what activities the platypus likes to do). Subsequently, the data of School A were scanned for participants who had used cancel inversion twice in the two contexts given in the picture story. Assuming that these learners would have used it a third time with an additional third context for cancel inversion, the criterion for the emergence of stage 6 would have been met. In that case, even more learners would possibly have produced the final stage in the data set. Hypothetically, this could have been the case for one learner in A1 (grade 1), nine in A2 (grade 4) and twelve in A3 (grade 4) (Schwirz, Maier, & Neubauer 2012).

In the second phase of the data collection (June 2012), the modified version of the picture story was piloted in School B. In that data set, 23 of the 38 participants used the third context for cancel inversion.

4.4.2 *Motivating tasks*

Within the IM settings (Schools A and B), a high level of motivation in working with the tasks was observed. All of the IM pupils seemed very engaged with the four tasks, which was in part directly expressed by several pupils who stated how much fun it was to work with the tasks.

With regard to the pupils from a traditional primary context (C1), however, an engagement with the tasks was not possible for some of the learners, as they lacked sufficient vocabulary to form a sentence or a question. As a result, long pauses punctuated the tasks. In one case, solving the storytelling task together with the learner was the only way to prevent him from refusing to finish the test. It became obvious that the restricted range of vocabulary limited the possibilities for some of the mainstream primary school pupils to solve the tasks. An example from child C 10 may illustrate this:

- (2) C 10: ehm (-) ehm (-) ((asks for the English term for “Schnabeltier”; IG helps)) (-) the platypus (-) ((asks for the English term for “Mülleimer”; IG helps)) the platypus (-) eh (-) ehm ((asks for the English term for “Mülleimer”, again; IG helps)) the platypus bin.

Other participants tried to overcome the problem of lacking vocabulary knowledge by making use of code-mixing. For instance, C 11 repeatedly forms sentences, like

- (3) C 11: *dann ist er in the bin reingeklettert* [then he climbed into the bin]
er hat a hole in the fence gemacht [he made a hole in the fence]

Obviously, sentences like these are not interpretable by means of PT, as only a few sentence constituents are expressed in the L2, whereas both the structure and the expression of the other words comply with the first language of the participant. This would indicate that the tasks did not appear motivating but rather intimidating to most of the pupils with lower L2 competence. In contrast to the IM pupils, who showed signs of disappointment when the test was over, many of the mainstream primary school pupils seemed to be relieved.

4.4.3 Comprehensibility

With regard to the comprehensibility of the tasks, there were also clear differences between the IM pupils and the mainstream primary school pupils (C1) (cf. H4.1). While the IM pupils mostly did not need any further explanations and could hardly wait for the IG to finish the instructions, the mainstream primary school pupils showed difficulties in working autonomously after the IG had explained the particular task. This lack of autonomy was also due to the restricted range of vocabulary. This became obvious when many of the pupils asked if they could, for example, tell the story in German first. While the German version was not problematic, the following attempts at telling the story in English were almost impossible without active support by the IG. It became obvious that it was not possible for the IG to play a rather passive role within the test situation. Instead, the IG became as active as the learners themselves, being responsible for translation or encouragement.

4.5 Discussion

With reference to the four research questions of the study, the discussion focuses on the grammatical language levels reached by the different groups, the influencing factors which were taken into consideration in this study, the suitability of the PT stages with respect to profiling IM pupils, and lastly, the suitability of the communicative tasks for working with children from different teaching programs, taking into account possible recommendations.

4.5.1 Results of the study

In general, the results show that the linguistic structures used by the learners comply with those proposed by Pienemann (1998). The implicational nature of the processing hierarchy was also confirmed. Nonetheless, a closer look at Tables 4–9 reveals that the results of some pupils include a ‘/’ (C03) or a ‘(+)’ (e.g. A15 and others), which might create the impression of representing counter-evidence for the implicational pattern. However, the (+) indicates that the structure is actually present in the data set, albeit not in the required number and variability of examples. This might be due to the child not making use of all the given contexts and should, therefore, not be considered as a counter-example for the implicational nature. With regard to C03, the ‘/’ indicates that the learner avoids the usage of the structure for the particular stage. Again, this does not contradict the implicational nature as it does not give any information about whether the structure has emerged in the interlanguage or not, but is rather due to the phenomenon that PT refers to as *learner variation* (Liebner & Pienemann 2011).

The results also indicate that the participants from the IM programs in grade 4 reach significantly higher stages than the pupils from traditional programs (Figure 4). The comparison among all first and among all third grades of IM programs also shows significant differences in the levels of attainment. These results suggest that there are differences between different IM schools and programs and even differences among classes within one school (B1 & B2). IM programs, in general, offer more L2 contact and a much higher intensity of L2 exposure. Additionally, from a longitudinal analysis of B1 and B2, in which two sets of tests were conducted a year apart, a strong correlation was shown between the increased amount of L2 classroom exposure and the improved language test scores. Another comparison with regard to intensity between A1 (IM) and C1 (traditional program) with a comparable L2 contact shows that the high intensity IM program leads to significantly higher results. This suggests that the higher the intensity and the duration of the L2 exposure, the higher the levels of grammatical competence that can be reached in a program.

Another difference among the programs might pertain to the fact that the IM teachers are native speakers or have a high command of English, which was not the case in group C1. More research would need to be conducted, however, to confirm this assumption.

Consideration should be given to the fact that this study works within a narrow field of comparison. As already mentioned in Section 2, comparing different programs of language teaching is difficult, due to various differences in the organization and the implementation of each particular program. According to Ellis (2008: 211), conclusions “need to be cautious”, as many social factors “interact among themselves, and their effect on learning depends to a large extent on the

setting”. A comprehensive comparison of these very different educational settings, however, is difficult, as it is almost impossible to consider all of the aspects which might influence the learner and, therefore, the results. Even within comparable educational settings, such as two IM schools, the results of the pupils may, for example, be influenced by factors such as the teaching methods or the personality of the teacher. A detailed comparison with an attempt to consider more of these influencing factors (e.g. age of onset, language use (L1 & L2) at home and at school, the quality of L2 input, socio-psychological factors, etc.) would go beyond the scope of this paper, but would be rewarding for future studies.

4.5.2 *Factors affecting L2 competence*

A statistical analysis of the effect size of different variables elicited through the parent’s questionnaire on the results confirmed the strong influence of L2 contact and L2 intensity. In this particular data set, L2 intensity has an effect size of almost 60% to explain the variance among the results, and represents an even stronger influence than L2 contact time with 36%. While sex, age and home language use do not show any effect, interestingly, prior experience with the target language, such as the attendance of a bilingual preschool or a long stay abroad, has a significant effect and explains roughly 10% of the variance. The majority of children with prior experience stem from an IM school (groups B1 & B2) and most of these children reached stages 5 and 6. If the data show a significant effect even if the variance among the stages reached is not especially high, it can be concluded that prior experience such as visiting a bilingual preschool actually does have a beneficial effect on L2 grammatical attainment in the long run. It would be interesting to corroborate these findings with regard to a larger database and other areas of linguistic competence.

4.5.3 *Suitability of PT for profiling IM learners*

The demonstrated ability of an IM pupil to reach high stages from an early grade reveals a so-called *ceiling effect* beginning in grade 2 (Schools A & B). This ceiling effect refers to an early emergence of structures at the highest stage (the ceiling), which makes it difficult to capture the development of these learners within the PT framework in a more detailed way. The modification of the picture story with an additional context for cancel inversion, the stage 6 structure, is supposed to have led to the emergence of more final stages in the data set. This is due to the participants’ use of the necessary three contexts in order for this structure to be considered as acquired according to the emergence criterion (cf. Section 4.3.1). With the modified picture story, a higher number of participants reached the highest stage possible than in the study with the original task, which added to the ceiling effect. This finding confirms H3, implying that more structures might be needed

to profile the development of advanced learners in more detail. A possible option would be a more fine-grained account of the variation within one stage¹³ or beyond stage 6. Currently, cancel inversion is the only structure included at stage 6, requiring the development of the most complex processing procedure described by the PT framework so far. If PT were able to account for more complex processing procedures which go beyond stage 6, the ceiling effect might be alleviated and thus render the framework more suitable to describe the development in later grades of intensive L2 programs. Structures such as the passive or certain types of subordinate clauses might be good candidates for such an endeavor.

An interesting phenomenon to be taken into account in this respect is the use of activity verbs instead of cognitive verbs to introduce indirect questions in the data (cf. Kersten 2009a), e.g.:

- (4) A 02: ... and there he looked where it lives.

Kersten (2009a: 276) points out that “the difference between indirect questions and other forms of subordination is not as clear-cut as theory would have it.” Arguably, learners seem to transfer “the inversion rule from indirect contexts to relative contexts” (Kersten 2009a: 276). Evidence for this phenomenon comes from Kersten (2009a: 275f, emphasis in the original):

- (5) 06.2: ... looks where the bees are. ... to see **where is the frog**.

- (6) 07.4: ... a *Markt* [market] **where can you buy Chinese things**

- (7) 08.3: ... the little frog who he caught. ... the little frog **who has he caught**

Based on such examples, Maier (2011: 18) argued that “[e]ven if the learner might not be able to differentiate the context of sub-clauses [and indirect questions] and over-use the rule, the application of cancel inversion is [...] evidence for the availability of the required procedures” (cf. Kersten 2009a). Issues such as these need to be taken into account if the PT framework were to be extended to relative clauses.

4.5.4 *Suitability of communicative tasks*

Within the IM settings, the data elicitation process worked well. The tasks elicited all the relevant structures and seemed to be motivating, enjoyable and comprehensible for all of the participants. With regard to the mainstream primary school context, however, some difficulties arose. This was especially true for the interview and the storytelling task, though less so for the habitual action task. The picture difference task posed the least difficulties. The game character of the latter seemed

13. For an approach to intra-stage development see Mansouri (2008).

a helpful incentive and even the less advanced learners seemed to enjoy it very much. With regard to the habitual action task, the vocabulary might have been easier for the pupils as it largely refers to their everyday life.

It is certainly true that, within a test situation, it is important for the tasks to also identify which structures are not yet used by a learner and that it is, therefore, unavoidable for the learners to face difficulties. Still, as the interviewers noticed it is essential that this issue should not result in frustration.

The study showed that, in contrast to advanced learners, learners from a mainstream primary school might need a certain amount of vocabulary help provided. If, for instance, the test would start with a matching task of English and German terms relevant for the interview, the necessary foundation could be laid in an entertaining and motivating way. The learner would then have the visual help of key vocabulary that s/he could make use of, which might facilitate the course of the different tasks to a great extent. However, it is also conceivable that when doing a task in which the translation from L1 to L2 is used this could favor the mainstream pupils, and hinder the IM pupil: the latter having become accustomed to using one language without switching back and forth between the L1 and the L2. Thus L1 activation might influence the interlanguage processes of both mainstream and IM pupils.¹⁴

With regard to H4.2, it became obvious that contexts for cancel inversion (stage 6) were present in the original version of the picture story, as participants produced this structure. This demonstrates that the pictures created for this context, showing an individual with thought bubbles (cf. Figure 2) in the process of thinking about something, serve the purpose of eliciting cancel inversion even for young learners. However, the learners did not produce the structure often enough. Here, the implementation of a third picture providing an additional context could remedy this issue.

Another problem of providing the learners with authentic contexts appeared with regard to the tenses. The use of the tense inflections was not stable throughout the data set. If no context is given, the use of both present and past tense inflections is justified (Kersten et al. 2002: 492). This raises the question of how to determine that a context in a narration is obligatory. Furthermore, the tense variation is due to the “cognitive-maturational development” of the learner (Kersten 2009a: 280) and thus cannot be expected from them. This holds especially true when considering that in oral communication and narration the skipping of tenses is natural even with native speakers. Still, “stable time reference is a specific competence

14. The implication of code-mixing on the learning process requires further investigation and cannot be accounted for in this article.

which has to be developed over time” (Kersten 2009a: 278). This was achieved, for instance, by A03, who consistently used the past tense in his narration.

In addition, the emergence of a linguistic structure such as, for example, 3rd ps. sg. -s, in a data set does not give any information (yet) about the form-function interface for the usage of that structure. Usually, the structure is analyzed as emerged when it fulfills the *grammatical* function of S-V agreement. This, however, does not indicate whether the *temporal* or *conceptual* function of the inflection is acquired by the learner. In other words, it is not possible to make any claims about whether the learner actually uses the inflection in order to indicate present tense or habitual action, or something entirely different. Competing theories on the distribution of functions of verbal inflections claim, for example, that verbal inflections are functionally distributed according to lexical aspect or according to foreground and background in narratives in early learner language (e.g. Andersen & Shirai 1994; Bardovi-Harlig 2000; Kersten 2009b). However, the habitual action task is supposed to elicit the 3rd. ps. sg. -s based on its conceptual function. It thus cannot be concluded that this function is actually understood by the learners.

Based on the previous argumentation, it is therefore questionable whether the task actually presents an obligatory context (in the literal sense of the word), or whether it has simply to be regarded as a context which facilitates the production of a specific grammatical structure independent of obligatory use. This distinction becomes relevant if differentiating between negative evidence (-) and avoidance (/) with regard to a specific structure (cf. Kersten 2009a). It is thus suggested to avoid the term *obligatory context* for contexts which elicit conceptual rather than grammatical functions.

5. Conclusion and future implications

This paper analyzed a number of studies on L2 production in different primary school settings. All studies confirmed the implicational nature of the stages proposed by PT. Learners from schools with IM programs clearly outperformed learners from a school with a traditional program, reaching significantly higher stages in their L2 production at the end of grade 4. These results also indicate that there are significant differences among the different IM programs.

A range of factors were taken into consideration in this study, some of which showed effects on the outcomes of IM and traditional programs. IM represents an EFL program with a much higher intensity and duration of L2 contact, starting in grade 1 and using the L2 as a means of communication in at least 50% of the curriculum. Statistical analysis confirmed the strong effect size of L2 intensity and L2 contact duration, with L2 intensity explaining almost twice as much variation in

the test results. In addition to this, an effect could also be found for prior experience with the L2, through, for example, the attendance of a bilingual preschool or a long stay abroad. This analysis shows the beneficial effect which a very early start with the L2 in preschool may have for children in the long run. No differences were found for sex, age and home language use, which means that children with a migration background in this data sample did not show any disadvantages in their L2 grammar production. However, as already stated (cf. Section 4.5.1), these factors only represent a selection of possible interrelated factors (Ellis 2008: 211). It is to be assumed that the consideration of additional factors, such as for example the quality of the language input, the degree of content-based teaching, the personality and the language competence of the teacher, etc., would be most rewarding in future studies.

The data of the IM learners showed a ceiling effect in their linguistic profiles. A majority of the learners reached stages 5 and 6 from grade 2 onwards, so that a more differentiated depiction of the progress in higher grades was difficult to obtain. These results thus reveal the strong impact that IM education has on the children's L2 language attainment. Tests for other linguistic competences or an extension of the stages predicted by PT, either intra-stage distinctions or beyond stage 6, would be very interesting to describe the high language competence of the pupils in a more detailed way in future analyses.

The communicative tasks used to elicit the data seemed to be successful especially for the advanced learners after a slight modification for cancel inversion, in that they provided enough contexts for all structures. An additional pilot test of the modified version with native speakers of English is recommendable.¹⁵ For less advanced learners, a modification of the test procedure is recommended, which includes the provision of key vocabulary and a more active role of the L1-speaking interviewer. A word of caution was voiced with regard to the term *obligatory context*, as it seems difficult to determine which contexts can actually be regarded as obligatory, taking the language learner's functionally restricted interlanguage system into account.

This paper focused on the comparison of the outcomes of different studies on the L2 acquisition in several primary school settings within the framework of PT. It is hoped that the results will provide a reference for future research with regard to the language levels achieved, the effect of different factors which are claimed to be beneficial for L2 development and, ultimately, the practical implications for teaching in different school programs.

15. We are grateful for feedback on this issue by participants of the 12th *International Symposium on Processability Approaches to Language Acquisition* 2012 in Ghent, Belgium.

References

- Andersen, R. W., & Shirai, Y. (1994). Discourse motivations for some cognitive acquisition principles. *Studies in Second Language Acquisition*, 16, 133–156.
doi:10.1017/S0272263100012845
- Bardovi-Harlig, K. (2000). *Tense and aspect in second language acquisition: Form, meaning, and use*. Oxford: Blackwell.
- Beckmann, U. (2006). Frühes Fremdsprachenlernen: Historischer Überblick. In M. Pienemann, J.-U. Kefßler, & E. Roos (Eds.), *Englischerwerb in der Grundschule* (pp. 11–23). Paderborn: Schöningh.
- Döpke, S. (1992). *One parent – One language. An interactional approach*. Amsterdam: John Benjamins. doi:10.1075/sibil.3
- Ellis, R. (1985). *Understanding second language acquisition*. Oxford: Oxford University Press.
- Ellis, R. (2008). *The study of second language acquisition*. Oxford: Oxford University Press.
- European Commission. (2003). *Promoting language learning and linguistic diversity: An action plan 2004 – 2006*. Accessed on 22 March 2013 from: (http://ec.europa.eu/education/doc/official/keydoc/actlang/act_lang_en.pdf)
- Eurydice. (2012). *Key data on teaching languages at school in Europe*. Brussels: Eurydice Unit. Accessed on 19 March 2013 from: (http://eacea.ec.europa.eu/education/eurydice/documents/key_data_series/143EN.pdf)
- Housen, A., Schoonjans, E., Janssens, S., Welcomme, A., & Pierrard, M. (2011). Conceptualizing and measuring the impact of contextual factors in instructed SLA – The role of language prominence. *International Journal of Applied Linguistics (IRAL)*, 49(2), 83–112.
- Johnston, M. (1985). *Syntactic and morphological progressions in learner English*. Canberra: Commonwealth Department of Immigration and Ethnic Affairs.
- Johnstone, R. (2002). *Addressing “The age factor”: Some implications for language policy*. Strasbourg: Council of Europe.
- Kersten, K., Imhoff, C., & Sauer, B. (2002). The acquisition of English verbs in an elementary school immersion program in Germany. In P. Burmeister, T. Piske, & A. Rohde (Eds.), *An integrated view of language development: Papers in honor of Henning Wode* (pp. 473–497). Trier: WVT.
- Kersten, K. (2009a). Profiling child ESL acquisition: Practical and methodological issues. In J.-U. Kefßler & D. Keatinge (Eds.), *Research in second language acquisition: Empirical evidence across languages* (pp. 267–294). Newcastle upon Tyne: Cambridge Scholars.
- Kersten, K. (2009b). *Verbal inflections in L2 child narratives: A study of lexical aspect & grounding*. Trier: WVT.
- Kersten, K. (2011). *Don't make so long, we waiting! Forschung und Praxis in bilingualen frühen Fremdsprachenlernen*. Mimeo: University of Hildesheim.
- Kefßler, J.-U., & Kohli, V. J. (2006). Erhebung von Sprachprofilen im frühen Englischerwerb: Kommunikative Tasks in Forschung und Fremdsprachenunterricht. In M. Pienemann, J.-U. Kefßler, & E. Roos (Eds.), *Englischerwerb in der Grundschule* (pp. 89–109). Paderborn: Schöningh.
- Kefßler, J.-U., & Liebner, M. (2011). Diagnosing L2 development. Rapid Profile. In M. Pienemann & J.-U. Kefßler, (Eds.), *Studying Processability Theory. An introductory textbook* (pp. 133–147). Amsterdam: John Benjamins. doi:10.1075/palart.1

- KMK. (2005). Kultusministerkonferenz: Bericht: Fremdsprachen in der Grundschule – Sachstand und Konzeption. Berlin 2004. Accessed on 19 March 2013 from: http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2005/2005_02_10-Fremdsp-Grundschule.pdf
- Kyriacou, C. (1997). *Effective teaching in schools: Theory and practice*. Cheltenham: Stanley Thornes.
- Larsen-Freeman, D., & Long, M. H. (1991). *An introduction to second language acquisition research*. New York, NY: Addison Wesley Longman.
- Leopold-Mudrack, A. (1998). *Fremdsprachenerziehung in der Primarstufe: Voraussetzungen, Konzept, Realisierung*. Berlin: Waxmann.
- Liebner, M., & Pienemann, M. (2011). Explaining learner variation. In M. Pienemann & J.-U. Kefler (Eds.), *Studying Processability Theory. An introductory textbook* (pp. 64–74). Amsterdam: John Benjamins. doi:10.1075/palart.1.05exp
- Lightbown, P.M., & Spada, N. (1999). *How languages are learned*. Oxford: Oxford University Press.
- Long, M. H. (1985). A role for instruction in second language acquisition: Task-based language teaching. In K. Hyltenstam & M. Pienemann (Eds.), *Modelling and assessing second language acquisition* (pp. 77–99). Clevedon: Multilingual Matters.
- Long, M. H. (2007). *Problems in SLA*. New York, NY: Lawrence Erlbaum Associates.
- Maier, E. (2011). *The acquisition of English as a second language in the primary classroom: Applying Processability Theory*. Unpublished BA thesis. University of Hildesheim.
- Mansouri, F. (2008). *Second language acquisition research: Theory-construction and testing*. Newcastle upon Tyne: Cambridge Scholars.
- Massler, U., Ioannou-Georgiou, S., & Steiert, C. (2010). CLIL teaching: Good practice. In S. Ioannou-Georgiou, & P. Pavlou (Eds.), *Guidelines for CLIL implementation in primary and pre-primary education* (pp. 67–98). Latsia: Cyprus Pedagogical Institute.
- Mehisto, P. (2011). Complexity competence: A tool for co-constructing CLIL programming. In D. Elsner & A. Wildemann (Eds.), *Sprachen Lernen – Sprachen Lehren/Language Learning – Language Teaching. Perspektiven für die Lehrerbildung in Europa/Prospects for Teacher Education across Europe* (pp. 121–140). Frankfurt: Peter Lang.
- Muñoz, C. (2011). Is input more significant than starting age in foreign language acquisition? *International Review of Applied Linguistics*, 49(2), 113–133.
- Neubauer, L. (2013). Assessing the grammatical language competence in primary school: A comparative study. Unpublished MEd thesis. University of Hildesheim.
- Niedersächsisches Kultusministerium. (2006). *Kerncurricula*. Accessed on 8 May 2013 from: <http://www.nibis.de/nibis.php?menid=3613>
- Pallotti, G. (2003). Cross-linguistic tests of Processability Theory: An operational definition of the emergence criterion. Presentation at EURASLA, Edinburgh, 18–21 September.
- Pallotti, G. (2007). An operational definition of the emergence criterion. *Applied Linguistics* 28(3), 361–382. doi:10.1093/applin/amm018
- Pienemann, M. (1998). *Language processing and second language development: Processability Theory*. Amsterdam: John Benjamins. doi:10.1075/sibil.15
- Pienemann, M. (2005). An introduction to Processability Theory. In M. Pienemann (Ed.), *Cross-Linguistic Aspects of Processability* (pp. 1–60). Amsterdam: John Benjamins. doi:10.1075/sibil.30.03pie

- Pienemann, M., & Kefler, J.-U. (2011a). *Rapid Profile*. Accessed on 22 March 2013 from: http://kw.uni-paderborn.de/fileadmin/kw/institute-einrichtungen/anglistikamerikanistik/English_Flyer.pdf
- Pienemann, M., & Kefler, J.-U. (2011b). *Rapid Profile 4 manual*. Accessed on 22 March 2013 from: http://kw.uni-paderborn.de/fileadmin/kw/institute-einrichtungen/anglistikamerikanistik/Rapid_Profile_Manual_Version4.pdf
- Pienemann, M., Kefler, J.-U., & Liebner, M. (2006). Englischerwerb in der Grundschule: Untersuchungsergebnisse im Überblick. In J.-U. Kefler & E. Roos (Eds.), *Englischerwerb in der Grundschule*, M. Pienemann (pp. 68–88). Paderborn: Schöningh.
- Pienemann, M., & Mackey, A. (1993). An empirical study of children's ESL development and rapid Profile. In P. McKay (Ed.), *ESL development: Language and literacy in schools* (pp. 115–259). Canberra: Commonwealth of Australia & National Languages and Literacy Institute of Australia.
- Piske, T., MacKay, I. R. A., & Flege, J. E. (2001). Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics*, 29, 191–215. doi:10.1006/jpho.2001.0134
- Piske, T. (2006). Zur Entwicklung der Englischkenntnisse bei deutschsprachigen Immersionsschülerinnen und -schülern im Grundschulalter. In N. Schlüter (Ed.), *Fortschritte im frühen Fremdsprachenlernen – Ausgewählte Tagungsbeiträge Weingarten 2004* (pp. 206–212). Berlin: Cornelsen.
- Roos, J. (2006). Frühes Fremdsprachenlernen: Eine Standortbestimmung. In M. Pienemann, J.-U. Kefler, & E. Roos (Eds.), *Englischerwerb in der Grundschule* (pp. 24–32). Paderborn: Schöningh.
- Schwirz, K. (2012). *Applying a second language acquisition theory to the EFL classroom: Processability Theory and its practicability for English teachers*. Unpublished MEd thesis. University of Hildesheim.
- Schwirz, K., Maier, E., & Neubauer, L. (2012). Applying Processability Theory to the primary immersion classroom. Presentation at the 12th International Symposium on Processability Approaches to Language Acquisition. Ghent, 28 September.
- Snow, M. A. (1990). Instructional methodology in immersion foreign language education. *Foreign Language Education: Issues and Strategies*, 113, 156–171.
- Weitz, M., Pahl, S., Flyman Mattsson, A., Buyl, A., & Kalbe, E. (2010). The Input Quality Observation Scheme (IQOS): The nature of L2 input and its influence on L2 development in bilingual preschools. In K. Kersten et al. (Eds.), *Bilingual Preschools, Vol. I: Learning and Development*, 5–44. Trier: WVT.
- Wenzel, L. (2011). *Second language learning as a process: Defining, applying and analyzing Processability Theory*. Unpublished BA thesis. University of Hildesheim.
- Wesche, M. B. (2002). Early French immersion: How has the original Canadian model stood the test of time? In P. Burmeister, T. Piske, & A. Rohde (Eds.), *An integrated view of language development* (pp. 357–379). Trier: WVT.
- Wiegand, J. (2012). *Immersive education: Effective second language learning in primary schools?* Unpublished BA thesis. University of Hildesheim.

Diagnosing L2-English in the communicative EFL Classroom

A task-based approach to individual and developmentally moderated focus on form in a meaning-focused setting

Jörg-U. Keßler & Mathias Liebner

Ludwigsburg University of Education / Carl-von-Ossietzky Gymnasium in Bonn

In this paper, we outline a solution to the problem that teachers for students in heterogeneous EFL classrooms need to provide lessons that enable language acquisition at different levels (here: stages of the PT Hierarchy). Therefore, we describe a learner- and learning-centred application of *Processability Theory* (henceforth PT; Pienemann 1998 and 2005) which covers a teaching unit that combines a communicative teaching approach with Second Language Acquisition diagnosis in order to foster L2 acquisition of individual learners in heterogeneous EFL classrooms. The focus here will be on how teachers can cope with heterogeneity in the classroom by offering suitable teaching units. We show how a combination of *Task-based Language Teaching* (e.g. Ellis 2003; Eckerth & Siepmann 2008) and PT provide the necessary theoretical framework for this teaching unit. Furthermore, *Rapid Profile* and the *Diagnostic Task Cycle* (Keßler 2008) are used within this overall framework for the teaching unit presented in this paper. This diagnostic approach should be seen as conceptual since it can be applied to various classroom settings. In the example presented here, students read a novel suitable for teenagers and produce *Podcasts* and thereby record natural communication in the classroom. This learner output delivers precise knowledge about second language development of each learner in a classroom. On this basis, the teacher can offer developmentally moderated treatment (e.g. Keßler 2008) and developmentally moderated focus on form (Di Biase 2008) to individual learners in heterogeneous EFL classrooms.

1. Introduction

Frequently students in a foreign language course do not acquire parts of the language to be learned although students are willing to learn and teachers prepare

their lessons carefully and with enthusiasm. Some reasons why this might be the case and an idea of how to tackle this issue in the classroom will be given in this chapter.

At first glance, foreign language courses and classrooms seem to consist of homogeneous learners. However, in schools all over the world, teaching and learning in heterogeneous classrooms is the norm rather than an exception. Keßler (2009) found evidence for a wide range of levels and developmental Interlanguage stages (cf. Pienemann 1998) in classrooms.

Therefore, language teachers face the challenge of planning courses that are learner-oriented and that cater to (individual) language acquisition of all participants. Mere textbook work with appropriate exercises often fails to achieve this. A possible solution to the problem is offered by output-oriented approaches like Task-based Language Teaching (cf. Keßler 2008; Vollmer 2008).

In this chapter we demonstrate how PT and Rapid Profile (Pienemann 1992 and 2006; Keßler 2006 and 2008; Keßler & Keatinge 2008; Keßler & Liebner 2012) can be used for a fast and valid diagnosis of the current level of acquisition in the language to be learned of each learner in a course. The results of the diagnosis deliver the framework for an application of Task-based Language Teaching (Ellis 2003; Eckerth & Siepmann 2008) and therefore have the potential to give individual treatment to all learners according to their stage of acquisition.

First, we briefly introduce Rapid Profile within the framework of PT (Pienemann 1998 and 2005; a detailed summary of Rapid Profile is provided by Keßler and Liebner 2011 in the PALART series). Then we describe a teaching unit set up for a secondary school, which consists of a diagnosis part with Rapid Profile and an individual treatment part. It is suitable for EFL learners in their fourth year and has been tested with eight grade students in Germany. To use a motivating approach, we developed a media and literature-based teaching unit with *Podcasts* and the novel 'Killing Mr. Griffin' by Lois Duncan.

All in all, we combine traditional literature teaching with new media and individual treatment in the EFL classroom, which leads to a unit that focuses on the needs of each language learner (cf. Keßler & Liebner 2012). It therefore focuses on suggestions and claims made by modern EFL curricula.

2. Profile Analysis with Rapid Profile

The Rapid Profile procedure is based on intensive research on natural and instructed second language acquisition (Pienemann 1992 and 2006; Keßler 2006; Keßler & Keatinge 2008; Keßler & Liebner 2012). We act on the assumption that language processing follows the same principles, because although there are

individual differences (e.g. short- and long-term memory; cf. Keßler & Plesser 2011), the architecture of the human brain is comparable among all learners (Pienemann 1998 and 2011).

Since the Rapid Profile procedure has its roots and background in PT we can ensure a valid and precise diagnosis of the level of language acquisition of a language learner. In this procedure, language learners are asked to work on communicative tasks. The language produced by them is analyzed according to the PT-Hierarchy in a computer-assisted procedure. After less than 20 minutes the teacher has a detailed profile of a language learner.

There are also other test procedures apart from profile analysis; many of them are based on the concept of proficiency. However, there is a “scope-precision-dilemma” (Pienemann & Keßler 2007) to be observed as proficiency-based approaches to language assessment try to cover and analyze as many aspects of a language as possible. This then leads to a lack of precision and validity. As profile analysis is very precise but lacks a wider scope we need to have a look at more aspects of language to yield a comprehensive picture of a language learner (cf. Grieshaber 2005:4). The last and all-encompassing step is an interaction between language acquisition research and foreign language pedagogy (cf. Keßler 2006). Illustration 1 reveals the whole picture, i.e. the roadmap to connect foreign language teaching in heterogeneous classrooms to language acquisition research and profile analysis with Rapid Profile.

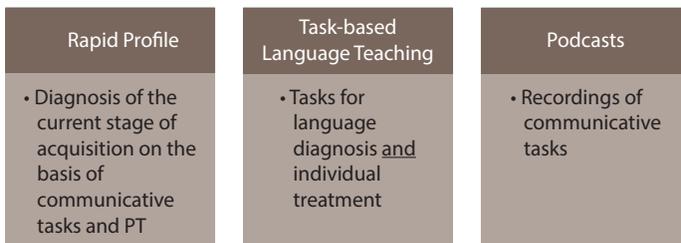


Illustration 1. Profile analysis in the EFL classroom

The Rapid Profile procedure has been described and tested in various studies (e.g. Pienemann 1992 and 2006; Keßler 2005 and 2006; Keßler 2008, Keßler, & Keatinge 2008; Keßler & Liebner 2011). Thus, we do not go into detail here. The focus in this paper is laid on individual treatment, which has been a major challenge in the EFL classroom. In order to guarantee that for each learner in a group, teachers compile media-based language profiles which reveal the precise stage of acquisition and help teachers to find supporting or challenging tasks that focus on learners’ needs.

Illustration 1 summarizes our media-based language diagnostic approach: The first component consists of the diagnosis of the current stage of acquisition of each learner. We show that this can be done in regular lessons. With the help of *Task-based Language Teaching* (e.g. Ellis 2003; Eckerth & Siepmann 2008), i.e. the second component of our model, tasks can be developed in such a way that they can be used both for the diagnosis as well as for the individual treatment in heterogeneous classrooms (cf. Keßler 2008). How these tasks can be used in our media-based diagnostic approach, will be further explained in the following section of this chapter. For the third component of our learner-based approach we use *Podcasts* to yield individual learner profiles and – on this basis – suggest individual treatment and developmentally moderated focus on form (Di Biase 2008) for oral production in EFL lessons.

3. Task-based Language Teaching within the teaching unit

The principle and underlying ideas of the task-based approach is well known and has been discussed widely (e.g. Ellis 2003; Eckerth & Siekmann 2008; Müller-Hartmann & Schocker-von-Dithfurt 2011; for a synopsis of various definitions of tasks and task-based language teaching see Ellis 2003 or Keßler & Plesser 2011). This is why we do not discuss the nature of task-based language teaching in detail in this chapter. We rather limit ourselves here to a brief introduction of how the task-based approach has been utilized within our teaching unit. We will demonstrate how tasks can successfully be administered both for Interlanguage diagnosis as well as for treatment.

A common feature of tasks in foreign language learning and in language acquisition research is the notion that authentic language use is present outside the classroom if there is a “gap” between interlocutors which needs to be filled (e.g. Keßler & Kohli 2006). These gaps can exist for various reasons, such as different previous knowledge of the speakers, different opinions between the interlocutors, or individual and cultural differences between them.

If we apply this knowledge to the EFL classroom, teachers should use activities in which such gaps are created on purpose (cf. Cook 1996). It is the learners’ task to fill those gaps with information and knowledge of their interlocutors by communicating and negotiation, i.e. *negotiation of meaning* (Long, 1996).

Within our teaching unit, we use the Task-based Language Teaching approach with a combination of “focus on meaning” (cf. Seedhouse 1997) and “focus on form” (cf. Gil 2002). Since both foci are present in the EFL classroom, we use tasks in the following sense:

As both modes are present in any classroom-based foreign language acquisition it would only be prudent to be a bit less preoccupied with either focusing on form or on meaning and rather join the two fields in a more holistic approach to foreign language pedagogy. Keßler (2008: 292)

Our approach presented in this chapter adds a diagnostic and a multi-medial component to the holistic approach mentioned in the quote and therefore adds a new dimension to learner-centered foreign language acquisition. In the past these components were not elegantly integrated, but were used as additional ideas and goals of EFL teaching and learning. The dual focus on content and – when necessary and developmentally appropriate – linguistic form as suggested in this chapter thus enables the learners not only to deal with content as precise as possible according to their individual current state of their L2 but also tells the teacher what kind of linguistic feedback the learners actually need and are ready to take in. By adding the diagnostic component to this classroom setting we support teachers not only to confidently know WHAT to teach and HOW but also – in terms of linguistic form – also WHEN to teach certain features to WHOM in the classroom. Instructed EFL teaching and learning thus becomes even more learner-centred and teachers are put in a position to cater to each individual learner in the classroom for her/his linguistic needs without having to provide totally different input to the learners. In other words: The learners can all work on the same task, however, each of them according to her/his own current state of Interlanguage development. Our unit thus provides a practical application of Di Biase's (2008) suggested developmentally-moderated focus on form.

4. Podcasts in the EFL classroom

The Web 2.0 has developed rapidly within the past decade. Weekly, new applications enter the market and millions of people share ideas and broadcast information with those tools. Beside social networks and blogs, also *Podcasts* have been given much attention recently and they are used more and more in schools. Apart from commercial offers and websites that help teachers to find ideas for the usage of *Podcasts* in their lessons, some schools also produce their own *Podcasts*.¹

The advantages can easily be named: With relatively low technical knowledge and low cost equipment, media-based EFL lessons can be planned and conducted easily. These lessons can be created learner-centered and filled with authentic

1. Some examples can be found here: <http://www.schoolpodcast.org/School/Listing> (22 February 2016).

communication. The *Podcasts* can be distributed via e-mail or other platforms and can then be used as a basis for the Interlanguage diagnosis of the EFL learners' language and as a starting point for the individual treatment of each learner (cf. Keßler & Liebner 2012).

This approach is exemplified by a teaching unit that draws from the novel "Killing Mr. Griffin" by Lois Duncan. In the following, section we will sketch out the unit and show how our concept can easily be adapted to many other teaching units.

5. The Teaching Unit – Individual treatment on a diagnostic basis

We would like to suggest a conceptual approach for a media-based profile analysis with Rapid Profile and individual treatment in heterogeneous learner groups. Exemplarily for this approach, we use the novel mentioned above. In this case, it is important to consider that we do not want to neglect the further analysis of the novel that has been read by the students and leave the spotlight on the students' environment and the plot of the novel. By doing so, the focus of the teaching unit fully remains on the content and the communicative power of the EFL classroom; any feedback and treatment of linguistic form is embedded into the discussion and analysis of the novel and comes as a "byproduct" – of course purposefully administered. The idea is to use the plot and encourage the students to create *Podcasts* with spontaneous speech production. Therefore, we see the potential of the teaching unit as a possibility to incorporate negotiation of meaning (cf. Section 3) by analyzing the novel further, diagnosing and working on the development of the learners' language skills and thus communicative and linguistic competence individually all at the same time within one teaching unit.

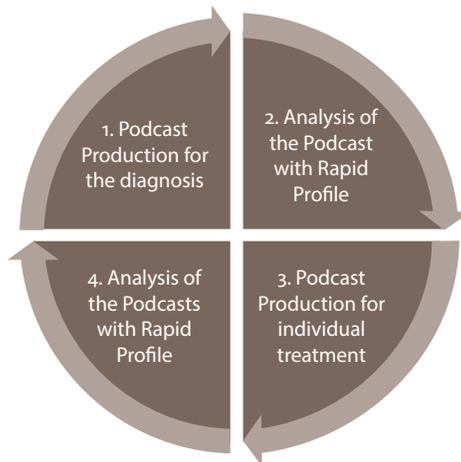


Illustration 2. Conceptual approach

Lesson	Subject	Commentary
1	How to Produce Your Own <i>Podcasts</i>	Introduction
2–3	Killing Mr. Griffin – What Happened Afterwards?	<i>Podcast</i> production for the diagnosis
4–5	Killing Mr. Griffin – What Happened Afterwards?	<i>Podcast</i> production for the discussion
6–7	Killing Mr. Griffin – A Closer Look at The Characters	<i>Podcast</i> production for individual treatment
8–9	Killing Mr. Griffin – A Closer Look at The Characters	<i>Podcast</i> production for the evaluation

Illustration 3. Unit plan

While Illustration 2 outlines the conceptual approach underlying the teaching unit described in this chapter, Illustration 3 summarizes the concept of the unit plan. Prerequisites for this teaching unit are that the students have already read the novel and that the analysis of its content has been dealt with in the preceding lessons. With the introduction, we ensure that students are able to produce *Podcasts* on their own after the first lesson. In lessons 2–5 students produce and present those *Podcasts* that are necessary for the diagnosis. Lessons 6–9 focus on individual treatment of each learner’s language. To do so, students produce *Podcasts* as well. In these, they work on the structures of the PT Hierarchy that need direct treatment according to the developmental stage that has been reached.

5.1 The diagnosis

The sequence described in this part of the unit focuses on the diagnosis. After the learners have been prepared for the *Podcast* production, the teacher introduces the proper task (cf. Illustration 4). Apart from the setting of the task the learners also receive role cards (cf. Illustration 5). At this time, the learners split up into groups of four and produce *Podcasts* according to their role cards. After having recorded the *Podcasts*, the learners hand them to the teacher.

The tasks were tested with grade 8 students in Germany who had – by the time of the project – attended EFL classes for about three and a half years. This knowledge, in combination with other Interlanguage studies (e.g. Pienemann, Keßler, & Liebner 2006; Keßler 2009), helped us to assume that these students would have at least reached stage 4 of the PT Hierarchy (Pienemann 1998). Thus, tasks were developed that encouraged learners to produce numerous stage 4 and stage 5 structures of the PT Hierarchy. Here we focused mainly on questions which can be asked according to copula questions and wh-copula questions such as “Is Peter at home?” or “Where is Peter?” (stage 4) and AUX-2nd-questions with or without negation such as “Where did you go afterwards?” or “Why didn’t you go home

*Worksheet 1 – Killing Mr. Griffin***What happened afterwards?****Task 1**

You read Lois Duncan’s “Killing Mr. Griffin”. The ending of the story was quite abrupt. We still do not know how the students were punished. Here is how the story might go on: A few days after the police arrested the students Susan and David are interviewed by the police.

1. Now it’s your turn.
 - Decide who is going to be Susan, David, Detective Baca and Detective Robinson.
 - Pick your role card and take ten minutes of preparation time. Try to find a solution for your situation. Do not take notes in advance or look at the other role cards.
2. Now press the record button and record your scene as a *Podcast*. Make sure that you have two rooms, one for Detective Baca and David and the other one for Detective Robinson and Susan. As in a real crime investigation Susan and David are not allowed to talk to each other after the interviews.
3. Once you are finished you have to record two more *Podcasts*:
 - David and Susan have the chance to talk about the police interviews and they want to find out about the other person’s interview. Please record the conversation.
 - The detectives have the task to find out whether Susan and David told the truth. Please record a conversation between the two detectives about the interviews.
4. Please upload your *Podcasts* to our platform.

Illustration 4. Worksheet 1

Role card I	Role card II
Susan McConnell Think about your situation after you were rescued and imagine a strategy that helps you and David to receive as little punishment as possible.	David Ruggles Think about your situation after you were arrested and imagine a strategy that helps receive no punishment at all.
Role Card III	Role Card IV
Detective Baca Develop a strategy with your partner that helps to force Susan and David to tell the truth about the murder.	Detective Robinson (Det. Baca’s partner) Develop a strategy with your partner that helps to force Susan and David to tell the truth about the murder.

Illustration 5. Role cards

after school?” (stage 5). The learners have also produced various other morpho-syntactical structures, which also helped to come to a complete analysis of the learner language.²

The tasks were carefully constructed so that the learners can produce spontaneous language. This has the effect that the learner language recorded in the *Podcasts* can be diagnosed with Rapid Profile (cf. Section 2). Therefore, the teacher has to analyze the *Podcasts* with Rapid Profile and the result is a diagnostic learner profile that offers a clear analysis of the developmental stage each learner has reached.

What is the advantage of this concept? The teacher can create learner-centered, media-based and motivating lessons in which learners produce authentic speech. At the same time the necessary data for the diagnosis is generated. Separate interviews with each student become obsolete and therefore this concept saves a lot of time by integrating the diagnosis into the lessons. Furthermore, the use of *Podcasts* offers a new stage of learner orientation and supports learner motivation as the learners in our project were not only keen to record their ideas but also liked to play them back to themselves and their peers. The *Podcasts* are recorded and sent to the teacher. This makes the interviews accessible for the teacher at any time. Added to that, these *Podcasts* are used in the following lessons for discussions on the various solutions of the learners. (cf. Illustration 6).

Task 2

Please present your *Podcasts* in class. Pay attention to the following criteria and take notes to the following questions:

- Did the actors create suspense?
- Was the performance believable?
- Were the actors creative?
- Do you see other possible improvements for their scene?

Illustration 6. Worksheet 2

5.2 Individual Treatment

Once the diagnosis with Rapid Profile is completed, the teacher can choose a suitable treatment from a task pool (cf. Illustration 7). Whereas task 2 (cf. Illustration 6) strictly focused on meaning, the Rapid Profile analysis provides

2. For a detailed description of profile analysis and task construction see (Kefßler 2008; Kefßler & Keatinge 2008; and Kefßler & Liebner 2011).

information on the learners' Interlanguage. Hence teachers can now offer a developmentally moderated focus on form (Di Biase 2008) and ask the students to work on tasks from the task pool and thereby individually work on tasks matched to their developmental stage in the PT-Hierarchy. The examples given here are based within the context of the novel and help to round up the picture. Again, we want to include the analysis of the novel into the diagnosis and the individual treatment (cf. Illustration 7). In order to do so, *Podcasts* are used once more, but this time not within a group of four, but rather individually. This has the big advantage that learners have more intensive language training.

Worksheet 3 – Killing Mr. Griffin

Task 3

Ask your teacher which task might be most suitable for you:

Task-Pool (A Closer Look at The Characters)

1. Record a *Podcast* in which you tell your classmates what could have happened to Mr. Griffin if
 - a. Mark had not been at that school.
 - b. Susan had told Mr. Griffin at their last meeting what the students planned.
2. Record a radio play as a *Podcast* in which you present one of the following situations:
 - a. You and your favorite character in the novel "Killing Mr. Griffin" have a talk. Your goal is to understand why the character acted as she/he did.
 - b. A classmate approaches you and asks if you help kidnapping your teacher.
3. Record a *Podcast* in which you play through an interview between
 - a. a newspaper writer and Mrs. Griffin.
 - b. a newspaper writer and Mark.
 - c. a newspaper writer and Susan.

Please note: You have to play both roles. You might want to change the tone of your voice.

Illustration 7. Worksheet 3

The tasks presented here focus on different morphosyntactical structures, which are assigned to different developmental stages within Processability Theory. If a teacher gives the third task of the task pool (cf. Illustration 7) to a learner, he or she will produce different question formation structures of stage 5 of the Processability-Hierarchy (Pienemann 1998). This task would be suitable if a learner has acquired structures of stage 4 according to the emergence criterion (Pienemann 1998) but has not acquired stage 5 yet.

A teacher now has the possibility to listen to the new *Podcasts* and check whether the treatment has been successful. Thus, the new *Podcasts* – which again were produced in lessons – are the new basis for a follow-up diagnosis of the learner language. This leads to a new choice of tasks for the individual treatment. In this way a natural diagnostic task cycle (cf. Keßler 2008) is created in the EFL classroom.

This is the last step of the teaching unit presented in this chapter. Apart from motivating EFL lessons for the learners, the teacher gained precise insight in the Interlanguage development of each learner. This can be used as the point of departure for further language teaching with individually provided developmentally moderated focus on form (Di Biase 2008) and eventually contribute to the creation of a developmentally moderated syllabus (cf. Keßler 2008). Additionally, the results of the Interlanguage diagnosis can be shared with the learners, parents or other teachers. In the case of Germany, this becomes more and more important, since a new school law in the state of North-Rhine Westphalia³ gives each student the right to receive individual feedback and treatment.

6. Conclusion

The teaching unit we have described here shows what an integration of literature, media-based lessons, Second Language Acquisition diagnosis, and individual treatment can look like in an EFL classroom and how it can be achieved by applying psycholinguistic knowledge to language teaching. We clarified that the diagnosis can become a regular and easily applicable part of the lessons and therefore avoid extra-curricular testing scenarios. Our concept presents an idea for modern EFL courses to the ongoing discussions regarding diagnosis and individual treatment. The concept presented here can easily be adapted to other topics and areas in the EFL classroom. A huge array of teaching units located in the task-based approach to language teaching and authentic communication are imaginable. Although we worked with grade 8 students in our project, basically all learner levels from beginners to intermediate can be taught according to this combined task-based approach on a diagnostic basis (cf. Keßler 2008; Pienemann & Keßler 2012).

Especially suitable for such teaching units are project-based courses in which learners are able to produce spontaneous, meaningful and authentic language.

3. The exact text can be found here: <http://www.schulministerium.nrw.de/docs/Recht/Schulrecht/Schulgesetz/Schulgesetz.pdf> (22 February 2016).

There is only one prerequisite: It has to be possible for a teacher to produce tasks within the teaching unit that offer the learner the chance to produce those structures within the Processability-Hierarchy necessary for the diagnosis and treatment of the learner language. As demonstrated in our paper this can easily be achieved within classroom settings.

References

- Cook, V. (1996). *Second language learning and second language teaching*. London: Arnold.
- Di Biase, B. (2008). Focus-on-form and development in L2 learning. In J.-U. Keßler (Ed.), *Processability approaches to second language development and second language learning*. Newcastle upon Tyne: Cambridge Scholars.
- Duncan, L. (1978). *Killing Mr. Griffin. Easy reader*. Britt Keson (Ed.). (2005). Egmont: Aschehough & Alinea.
- Eckerth, J. (2008). Task-based language learning and teaching – old wine in new bottles? In Eckerth & Siepmann (Eds.), pp. 13–46.
- Eckerth, J., & Siepmann, S. (Eds.). (2008). *Research on task-based language learning and teaching. Theoretical, methodological and pedagogical perspectives*. Frankfurt: Peter Lang.
- Ellis, R. (2003). *Task-based language learning and teaching*. Oxford: Oxford University Press.
- Gil, G. (2002). Two complementary modes of foreign language classroom interaction. *EFL Journal*, 56(3), 273–279.
- Grießhaber, W. (2005). Sprachstandsdiagnose im Zweitspracherwerb: Funktional-pragmatische Fundierung der Profilanalyse. Accessed on 6 March 2009 from: <http://spzwww.uni-muenster.de/~griesha/pub/tprofilanalyse-azm-05.pdf>
- Keßler, J.-U. (2005). Fachdidaktik meets Psycholinguistik – Heterogenität im Englischunterricht erkennen, verstehen und als Chance nutzen, In K. Bräu & U. Schwerdt (Eds.), *Heterogenität als Chance. Vom Umgang mit Gleichheit und Differenz in der Schule* (pp. 263–284). Münster: LIT; S.
- Keßler, J.-U. (2006). Englischerwerb im Anfangsunterricht diagnostizieren. *Linguistische Profilanalysen am Übergang von der Primarstufe in die Sekundarstufe I*. Tübingen: Narr.
- Keßler, J.-U. (2008). Communicative tasks and second language profiling: Linguistic and pedagogical implications. In Eckerth & Siepmann (Eds.), pp. 291–310.
- Keßler, J.-U. (2009). Englischdidaktik in ‚Erklärungsnot‘: Implizites und explizites Wissen und die Rolle der Bewusstmachung im schulischen Englischerwerb. In R. Vogt (Eds.), *Erklären: Gesprächsanalytische und fachdidaktische Perspektiven* (pp. 93–107). Tübingen: Stauffenberg.
- Keßler, J.-U., & Keatinge, D. (2008). Profiling oral second language development. In J.-U. Keßler (Eds.), *Processability approaches to second language development and second language learning* (pp. 167–197). Newcastle upon Tyne: Cambridge Scholars.
- Keßler, J.-U., & Kohli, V. J. (2006). Erhebung von Sprachprofilen im frühen Englischerwerb: Kommunikative Tasks in Forschung und Fremdsprachenunterricht. In M. Pienemann, J.-U. Keßler, & E. Roos (Eds.), *Englischerwerb in der Grundschule. Ein Studien- und Arbeitsbuch* (pp. 89–96). Paderborn: Schöningh/UTB.

- Kefler, J.-U., & Liebner, M. (2011). Diagnosing L2 development: Rapid Profile. In M. Pienemann & J.-U. Kefler (Eds.), *Studying Processability Theory. An introductory textbook* (pp. 133–147). Amsterdam: John Benjamins. doi:10.1075/palart.1
- Kefler, J.-U., & Liebner, M. (2012). Medienbasierte Lernstandsdiagnostik und individuelle Lernerförderung – Neokommunikativer Fremdsprachenunterricht in heterogenen Lerngruppen. In M. Reinfried & L. Volkmann (Eds.), *Medien im neokommunikativen Fremdsprachenunterricht. Einsatzformen, Inhalte, Lernerkompetenzen* (pp. 303–317). Frankfurt: Peter Lang.
- Kefler, J.-U., & Plesser, A. (2011). *Teaching grammar*. Paderborn: Schöningh/UTB.
- Lightbown, P., Spada, N. (2006). *How languages are learned* (3rd rev. ed.). Oxford: Oxford University Press.
- Long, M.H. (1996). The role of linguistic environment in second language acquisition. In W. Ritchie & T. Bhatia (Eds.), *Handbook of research on second language acquisition* (pp. 413–468). New York, NY: Academic Press.
- Müller-Hartmann, A., & Schocker-von-Ditfurth, M. (2011). *Teaching English: Task-supported language learning*. Paderborn: Schöningh/UTB.
- Pienemann, M. (1992). *Assessing second language acquisition through Rapid Profile*. Ms. Sydney.
- Pienemann, M. (1998). *Language processing and second language development: Processability Theory*. Amsterdam: John Benjamins. doi:10.1075/sibil.15
- Pienemann, M. (2005). *Cross-linguistic aspects of Processability Theory*. Amsterdam: John Benjamins. doi:10.1075/sibil.30
- Pienemann, M. (2006). Spracherwerb in der Schule. Was in den Köpfen der Kinder vorgeht. In M. Pienemann, J.-U. Kefler, & E. Roos (Eds.), *Englischerwerb in der Grundschule. Ein Studien- und Arbeitsbuch* (pp. 33–63). Paderborn: Schöningh/UTB.
- Pienemann, M. (2011). The psycholinguistic basis of PT. In M. Pienemann & J.-U. Kefler (Eds.), *Studying Processability Theory. An introductory textbook* (pp. 27–49). Amsterdam: John Benjamins. doi:10.1075/palart.1
- Pienemann, M., & Kefler, J.-U. (2007). Measuring bilingualism. In P. Auer & L. Wei (Eds.), *Handbook of applied linguistics, Vol. 5: Multilingualism and multilingual communication* (pp. 247–275). Berlin: Mouton/de Gruyter.
- Pienemann, M., & Kefler, J.-U. (2012). Processability Theory, In S. Gass & A. Mackey (Eds.), *Handbook of second language acquisition* (pp. 228–247). New York, NY: Routledge.
- Pienemann, M., Kefler, J.-U., & Liebner, M. (2006). Englischerwerb in der Grundschule: Untersuchungsergebnisse im Überblick. In M. Pienemann, J.-U. Kefler, & E. Roos (Eds.), *Englischerwerb in der Grundschule. Ein Studien- und Arbeitsbuch* (pp. 67–88). Paderborn: Schöningh/UTB.
- Seedhouse, P. (1997). Combining form and meaning. *ELT Journal*, 5(14), 336–344. doi:10.1093/elt/51.4.336
- Vollmer, H.J. (2008). Constructing tasks for content and language integrated learning and assessment. In Eckerth & Siepmann (Eds.), (pp. 227–290)

The cognitive processes elicited by L2 listening test tasks – A validation study

Henning Rossa

TU Dortmund University

This paper is concerned with an investigation into the validity of a listening comprehension test that was developed for a large-scale assessment project. The study draws on qualitative data, employing a think-aloud technique and stimulated recall interviews. The informants (n=18) were purposefully and randomly sampled from a group (n=121) of year 9 learners (ages 14–16) of English as a foreign language (EFL) in German schools. Subjects were asked to think aloud while they were solving the multiple choice-items of the listening test. Construct-relevant and -irrelevant processes were identified and analysed with regard to their distribution across the two subsamples and their relative contribution to correct item responses. The results provide validity evidence for the listening tests in general. A few test items, however, were shown to elicit test-taking processes and strategies that compromise the measurement outcomes.¹

1. Introduction

This paper is based on a validation study that investigates the processes and strategies German EFL-learners employ in their attempts at solving multiple-choice L2 listening comprehension test tasks. The test tasks were developed and used in the context of DESI, a large-scale assessment study that aimed at describing the language abilities (German and English)² of year nine students in German schools (cf. Nold & Rossa 2007). Data on learners' progress in developing their foreign language abilities during their ninth year of school education was analyzed in connection with variables pertaining to the educational context and characteristics of the type of L2 instruction learners were exposed to – as assessed by a videography study (cf. DESI-Konsortium 2008).

-
1. The study was previously published in more detail in Rossa 2012.
 2. For the vast majority of these learners German is the L1, while English is an L2, acquired as a foreign language.

The study presented here seeks to provide insights into two particularly thorny issues of language testing research: defining the construct that is to be measured and checking the validity of the measurement procedure. These substantive issues are positioned at an interface between SLA research and language testing research, as identified by Brindley (1998: 127 ff.). Pienemann and Keßler define this interface as the point where studies on language development profiling and proficiency testing have to come to terms with the scope of their research and the precision of their measurement instruments. These reference points for research in Applied Linguistics currently seem to contradict each other (cf. Pienemann & Keßler 2007: 257). The objective of the present study, then, is to demonstrate how this dilemma could be resolved, by focusing both on the nature of the construct that is being measured (scope) and the validity of the measurement instruments (precision).

In contrast to most contributions to this volume, this is, essentially, not a study of interlanguage development, but rather an empirical investigation situated at the crossroads of language testing and foreign language education research. The study's cognitive approach to understanding test task performance as evidence of the learner's underlying listening ability does, however, share some of the basic theoretical propositions of Processability Theory and subscribes to its underlying logic: "[A]t any stage of development the learner can produce and comprehend only those L2 linguistic forms which the current state of the language processor can manage" (Pienemann 2003: 686).

The observation that L2 learners move from emergence (of a form) to its mastery in producing the L2, then, may also be true for the reversed processes of language reception, where learners move from partial comprehension (of a concept expressed in L2 forms) to full comprehension. The notion of incomplete or 'emerging' comprehension is of particular relevance for this study, which seeks to understand the processing that precedes successful and unsuccessful item responses on a listening test. A key concept that can account for basic comprehension processes is the idea of incremental and parallel processing of grammatical and propositional information (cf. Levelt 1989), which also features in the PT architecture.

Since the test construct and the theoretical approach employed in the present study depict comprehension as a complex mental process, in which the language user orchestrates various cognitive and metacognitive resources, syntactic processing of (isolated) linguistic forms, which would be of interest for studies on comprehension within the PT framework (e.g. Senécal 2011), can only theoretically be acknowledged as a central mechanism to allow L2 comprehension and as one possible source of failed comprehension or misunderstanding. Similar to PT-based research, the present study also acknowledges the psychological

constraints of L2 processing, e.g. the automatic nature of grammatical processing, the limited capacity of procedural memory and lexical access as a parallel process (cf. Pienemann 1998: 56–66).

Both concepts, listening ability and test validity, share a somewhat paradoxical position in Applied Linguistics. Their importance is stressed in numerous theoretical publications, but this has not resulted in equally strong empirical efforts. Samuel Messick, whose seminal work is the foundation of current validity theory in language testing research, laments: “Many test makers acknowledge a responsibility for providing general validity evidence of the instrumental value of a test, but very few actually do it” (Messick 1992: 89).

In more recent publications Weir confirms Messick’s analysis (Weir 2005: 11), while Bachman and Alderson describe a similar dilemma in the preface to Buck’s reference book on assessing listening: “The assessment of listening abilities is one of the least understood, least developed and yet one of the most important areas of language testing and assessment” (Buck 2001: X).

The relative scarcity of empirical research on the assessment of listening abilities and test validity seems to be due to the fundamental characteristics of these constructs. Models of listening currently map it as a complex interactive process that is largely based on automatised mental operations, thus rendering it nearly unobservable. Validity, on the other hand, receives continuing attention from theoretical and conceptual work in psychology, educational measurement and language testing research. It has evolved into a concept that refers to the measurement procedure, the interpretation of test scores and the social dimensions of testing: impact and washback (cf. Bachman 2004, 2005; Chapelle 1998; Kane 2001; Kunnan 2000; Messick 1989, 1996; Mislevy 1996).

The proponents of such an extended definition of validity admit that it may discourage many working researchers: “Understanding the social function of tests can be seen by many authors as introducing an unmanageable aspect into language testing research, opening a Pandora’s box of issues with no chance of practical resolution” (McNamara & Roever 2006: 40–41). Due to the complexity of the concept validation studies are currently forced to select a specific facet of validity for investigation and consider the growing number of statistical, psychometric, qualitative and introspective methods for data elicitation and analysis accordingly (cf. Stoyhoff 2009: 34). The validity arguments collected in such research contribute to our understanding of three different levels of validity:

1. Task development, test administration, test-taker responses
2. Interpretation of test scores
3. Social consequences of the interpretation of test scores

In spite of the multifaceted nature of validity most authors agree that the core of the concept, often labelled as “construct validity”, deals with the question: Do the test tasks actually measure what the test is supposed to measure? (Cronbach & Meehl 1955; Messick 1989; Bachman 2004). Construct validity studies, however, tend to neglect the first level of validity arguments described above and focus on the second level, the interpretation of test scores, instead. Such studies are based on theoretically postulated relationships between the test scores in question and scores related to other ability constructs. The dominating role of these studies has recently been challenged in psychological testing (cf. Borsboom, Cramer, Kievit, Zand Scholten, & Franic 2009; Friese & Fiedler 2010). Borsboom et al. question the value of correlational research arguments that are built upon statistical relationships between constructs in a nomological network. They argue that these relationships cannot be tested empirically and do not reflect the essential issue of validity (cf. Borsboom et al. 2009: 166).

In an earlier proposal motivated by the insufficient applicability of construct validity to empirical research Borsboom et al. developed an alternative concept they define as ‘test validity’ (Borsboom, van Heerden, & Mellenbergh 2004). It is based on the assumption that variations in the measurement outcome are not only correlated to but caused by variations of the construct being measured. This realist perspective on measurement implies that validation studies should adopt a research perspective that differs from the mainstream of current validation practices:

What needs to be tested is not a theory about the relation between the attribute measured and other attributes but a theory of response behavior. Somewhere in the chain of events that occurs between item administration and item response, the measured attribute must play a causal role in determining what value the measurements outcomes will take; otherwise, the test cannot be valid for measuring the attribute. (Borsboom, van Heerden, & Mellenbergh 2004: 1062)

It follows from the logic of this argument that “the locus of evidence for validity” (ibid.) can be found on the level of test-takers’ interactions with the tasks. These processes reveal to what extent the test-developers were successful in translating the construct into task demands. In the context of language testing this approach to validation is mirrored in Weir’s concept of ‘theory-based validity’. It consists of arguments which demonstrate a match between processes theoretically associated with situations of language use and actual task-processing in the language test situation (cf. Weir 2005: 18).

The present study adopts ‘validity’ as the term that represents its central concern, drawing on Borsboom et al.’s general concept of ‘test validity’, Weir’s notion of ‘theory-based validity’ and the fundamental assumption underlying the logic of language testing: If test-takers perform successfully on a given test item, this

provides a piece of evidence that suggests they have acquired the facets of language ability the item reflects. Thus, the central issue for validation research is the question to what extent test-takers succeed and fail on the test items for reasons relevant to the construct. Accordingly, validity is defined in the context of this study as follows: A language test is valid for measuring the attribute specified in the construct, if (a) variations in the attribute produce variations in the measurement outcome and (b) the processes that coincide with successful item responses match the processes of language use specified in the construct.

2. Methods

Based on the definition of validity given above, the study addresses the following research questions:

1. What are the mental processes test-takers engage in while they attempt to solve the DESI EFL listening comprehension test items?
2. To what extent do the mental processes of the test-takers correspond with facets of the EFL listening comprehension construct specified for the test items?
3. What is the nature of test-takers' mental processing that coincides with correct and incorrect item responses?

These research questions yield two objectives for the present study: Firstly, it is necessary to collect information about the processing test-takers engage in while they are trying to solve the test items, and secondly, these mental operations must be compared with the processes underlying the construct specified for the test tasks.

Consequently, the central theoretical frame of reference for this study consists of the DESI EFL listening construct and the research on L2 listening comprehension processes and strategies that informed its development (Buck 1991, 1992, 2001; Ross 1997; Kintsch 1998; Rost 2002; Vandergrift 2003). According to the construct the test tasks seek to measure the following facets of listening ability:

- Processing short and extended samples of spoken language (English [Near-RP and General Canadian], authentic speech rates, generally clear articulation, scripted texts) in real time.
- Understanding the linguistic information that is presented on the local level of the input text (understanding details)
- Connecting pieces of information in order to develop a mental model which allows comprehension on the global levels of the input text (understanding gist)

- Matching explicitly and implicitly presented information (actions, emotions, intentions) with language knowledge and background knowledge to recognize and retrieve, to infer, and to interpret this information.
- Constructing a representation of information presented in the aural mode that allows the listener to understand paraphrases of that information in other (written) contexts. (Nold, Rossa & Hartig 2008:99)

A process-oriented approach to test validation can additionally capture the perspectives the test-takers may have to offer with regard to the ways in which they experience the test situation. As a consequence, methodological suggestions from previous studies on test-takers' perceptions in language testing research were considered in the process of selecting appropriate methods for data elicitation and analysis (cf. Cohen 2000, 2007; Shohamy 2001).

2.1 Integrating qualitative and quantitative data in a mixed methods approach

The individual perspectives of the test-takers and the variability and complexity of their mental processing suggest a qualitative approach to research design. However, the large-scale assessment context of the study supplies additional quantitative perspectives on the validity of the test tasks, markedly data on the psychometric qualities of the test items and a system of task characteristics that was developed to predict item difficulty parameters. The research design of the present study thus implements a mixed methods approach that Creswell et al. define as "concurrent triangulation" (Creswell, Plano Clark, Gutmann, & Hanson 2007:224). The study builds on qualitative data on the test-takers' mental operations related to test task processing and integrates quantitative data on the characteristics of the test tasks in the phase of data analysis and interpretation (cf. Nold, Rossa & Hartig 2008). Additionally, relationships among the phenomena identified in qualitative data analysis are subjected to the scrutiny of statistical tests.

This approach serves two main purposes: First, it is hoped that integrating data from different methodological perspectives will result in a more comprehensive account of the objects of research (cf. Denzin 1970: 300ff.). Second, confirming results of qualitative research with the help of quantitative methods of data analysis may enhance "the integrity of the findings" (Bryman 2006: 106). This objective makes use of one of the earliest ideas associated with the concept of triangulation Campbell and Fiske put forth in the context of a theory of psychological testing. They argue that the validity of research results can be improved by applying multiple methods of measurement (cf. Campbell & Fiske 1959). In a similar vein, Webb et al. claim in an early reference to the concept of triangulation in social

science research that the “most persuasive evidence comes through a triangulation of measurement processes” (Webb, Campbell, Schwartz, & Sechrest 1962: 3).

A mixed methods research design has to be based on explicit epistemological decisions which would otherwise be clearly delineated in an approach that is either qualitative or quantitative. One such issue is the role of hypotheses and theoretical frameworks in the research design. While hypotheses are central to quantitative designs, they seem to be incompatible with inductive reasoning in qualitative research, as they might manipulate or limit the exploratory scope of the study (cf. Maxwell 2005: 70).

From the pragmatist point of view of a mixed methods approach, however, it becomes apparent that a priori statements about the phenomenon in question, which draw on previous research efforts, may provide a sense of direction and structure for the processes of data elicitation and analysis. Additionally, for researchers who cannot (or choose not to) adopt the role of a pure observer, as is the case in this study, the process of making explicit the theoretical assumptions and previous knowledge the study builds on helps contextualise the ensuing research results. Handbooks on qualitative research methodology (e.g. Denzin & Lincoln 2000; Maxwell 2005; Patton 2002; Richards 2003; Smith 2003; Yin 2003) also suggest that researchers should be aware of theoretical categories which may help describe the research objects before they start collecting data:

Observers do not enter the field with a completely blank slate. While the inductive nature of qualitative inquiry emphasizes the importance of being open to whatever one can learn, some way of organizing the complexity of experience is virtually a prerequisite for perception itself. (Patton 2002: 279)

These conceptual entities feature in qualitative research as ‘sensitizing concepts’ (Blumer 1954) or theoretical ‘propositions’ (Miles & Huberman 2009: 75) which make up the ‘conceptual framework’ (Maxwell 2005: 33) of qualitative designs. In the processes of data analysis and interpretation these theoretical propositions may be extended or revised (cf. *ibid.*: 70).

2.2 Conceptual framework: Theoretical propositions

2.2.1 *Theoretical propositions concerning the first research question*

Test-takers employ cognitive processes relevant to language use while they try to comprehend the listening texts and solve the respective test tasks. Asking test-takers to verbalise their thoughts while they listen to the text will inevitably result in cognitive overload, so data can only be collected in the subsequent phase of task-processing. Consequently, the verbal reports of the test-takers will mainly provide evidence for the results of cognitive processing that are available for verbalisation,

such as text information recalled in working memory. Additionally, the verbal reports will probably provide insights into inferential reasoning processes related to the matching of information recalled with the options of the multiple-choice items.

Test-takers may monitor and evaluate their comprehension and task-processing with the help of metacognitive and affective strategies. These are likely to resemble strategies of language use, learning strategies, comprehension strategies and test-taking strategies as previously identified in cognitive psychology, applied linguistics and language testing research.

2.2.2 *Theoretical propositions concerning the second research question*

According to the definition of ‘validity’ adopted in this study, data analysis should focus on the extent to which test-takers are actually engaged in language-use processes that are specified in the test construct while they respond to test items. Messick’s unified account of validity provides two categories that refer to the possible threats to validity implied in the focus of this study: construct-underrepresentation and construct-irrelevant variance. Construct-underrepresentation exists, if the test tasks do not sufficiently reflect the facets of language ability specified in the construct. This means that the present study will have to assess to what extent the test-taking processes that become evident in the verbal reports cover all the facets of the test construct. Construct-irrelevant variance can occur, if test formats demand mental operations from the test-takers that are not specified in the construct and if the ability to execute these construct-irrelevant processes is variably distributed among test-takers. When construct-irrelevant task demands influence item responses and, ultimately, success on the test, this will call into question the validity of test-score interpretations (cf. Kane 2001: 333).

2.2.3 *Theoretical propositions concerning the third research question*

The results of language ability measurements in DESI were scaled within a probabilistic measurement model (Rasch-model, Rasch 1960) based on item response theory (cf. Lazarsfeld 1960; Lord 1980; Rasch 1961; Rasch 1968). The model assumes the existence of a stochastic relationship between item responses (succeeding or failing to solve a given task), a latent trait and task demands:

A person having greater ability than another should have the greater probability of solving any item of the type in question, and similarly, one item being more difficult than another one means that for any person the probability of solving the second item correctly is the greater one. (Rasch 1980: 117)

Within the validity framework adopted in this study the latent trait of the test-takers is revealed in the mental operations that aim at successfully attending to

task demands. The stochastic relationship between test-taker abilities and item responses can only be assumed, if successful item responses are attended by construct-relevant processing. The sole use of construct-irrelevant processes that might be elicited by characteristics of the test asks should ideally not contribute to successful task performance, or, worse still, interfere with relevant processes in a way that will impede successful item responses.

The study has made use of a purposeful random sampling of extreme cases to select informants for the collection of verbal data on test-taking processes (cf. Patton 2002: 169). This approach was chosen based on the expectation that more light could be shed on the variable mental operations elicited by test tasks by contrasting test-takers on different levels of their L2 development. The original sample (N=121) matched the target population of the DESI-study with regard to the most fundamental variables. The sample was made up of German pupils in year nine classes from four different types of secondary schools. At the time of data collection, they had learned English as a foreign language for five years. Of all the learners, 23 had received EFL instruction in a late partial immersion programme which included two additional hours of English teaching per week for two years and three subjects (Geography, Social Sciences and History), where English was the language of instruction.

The original sample of informants took the DESI test-modules 'reading comprehension', 'listening comprehension' and 'text reconstruction' (C-test) and responded to questionnaire items concerning their motivation to learn English at school and their use of language learning strategies (cf. DESI Konsortium 2008).

The C-test scores of the original sample captured the entire spectrum of language development (scores are below level 1 and go up to the highest level 5) as identified in the main study of the DESI project. In a second phase of sampling, two extreme groups were selected. They included test-takers whose C-test scores were particularly low (15–35 correct responses on a 100 item test) or high (75–95 correct responses). Despite the ongoing debate about the exact nature of the construct the C-test intends to measure, its generally strong psychometric qualities and consistently high correlation indices with a number of relevant L2 language skill areas do mark the C-test as a useful sampling variable that provides an estimate of a test taker's general language proficiency (cf. Grotjahn & Eckes 2006).

The extreme subsamples were then selected by random sampling from the two score bands identified above. These two random samples were controlled by three criteria. First, the mean scores of the two groups should be of equal distance to the mean of the original sample. Second, the distributions of the two subsamples should be similar to the skewness and kurtosis measures of the original sample. Third, the distribution of the gender variable, which is characteristic of both score bands (67% female, 33% male), should be maintained in the subsamples. These

three criteria were chosen to preserve some of the central characteristics of the original sample in the reduced sample for the qualitative study.

Table 1 presents the results of the phases of purposeful random sampling concerning the attempt to maintain the distribution parameters of the original sample within the constraints of the small sample sizes for the collection of verbal data.

Table 1. Descriptive statistics concerning the distributions of the subsamples

Sample	N	M	SD	skewness	curtosis
Original sample	121	52,60	21,4	-0,20	-0,80
Score bands LO	16	23,81	5,12	-0,22	-1,21
Randomised subsample LO	9	26,89	9,70	0,78	-0,26
Score bands HI	26	80,81	5,48	0,34	-0,79
Randomised subsample HI	9	83,00	5,74	0,61	-0,45

The mean score of subsample LO is 1.20 standard deviations below the mean of the original sample, while the mean score of subsample HI is 1.42 standard deviations above the original mean score.

2.3 Collecting verbal data on test-taking processes

The present study has applied the think-aloud method to collect introspective verbal reports on the test-takers' cognitive processing, drawing on the suggestions put forth in other studies that focus on learner cognition (Ericsson 2003; Ericsson & Simon 1993; Haastrup 1987; Van Someren, Barnard, & Sandberg 1994). Retrospective 'stimulated recall' interviews (Gass & Mackey 2000) were used to enhance the analysis of the think-aloud data and investigate the individual perspectives of the test-takers.

The appropriacy of think-aloud data for the study of cognitive processes rests on a simple information-processing model of human cognition: a substantial amount of our thoughts can (a) be kept in working memory for some time, and (b) are available for verbalisation (cf. Ericsson & Simon 1993). These theoretical assumptions seem to fit the theoretical framework of the present study, mainly because they are compatible with the construction-integration model of text comprehension (Kintsch & van Dijk 1978; Kintsch 1998) that is central to the DESI EFL listening comprehension construct. Kintsch and Ericsson have also worked together to explain comprehension processes based on the concept of 'long-term working memory', a part of memory that provides access to long-term memory for working memory for the construction of mental text representations (Kintsch, Patel, & Ericsson 1999). According to the information-processing model

underlying Ericsson and Simon's work on the think-aloud method, the cognitions exchanged between working memory and long-term memory should also be available for verbalisation.

Think-aloud data were recorded in the present study to tap the cognitive processes related to taking the DESI listening test and to capture the text propositions test-takers integrate into their mental representations of the input texts and match with the propositions provided in the answer options of the multiple-choice items.

The informants received individual training sessions that aimed at familiarising them with the think-aloud method. The training included board games and problem-solving tasks and introduced standardised instructions of the method based on Ericsson and Simon (1993) and van Someren et al. (1994): "Please tell me everything that goes through your head, no matter how unimportant it may seem. When you read something, please read it aloud." Instructions were given in German, but informants were told that they were free to choose which language to use for verbalisation. All informants but one, a native speaker of Dari from Afghanistan, who thought aloud in English, chose to verbalise their thoughts in German, which is the L1 for fourteen and an L2 for four of the informants.

During the phase of data elicitation the instructions were reduced to encouraging the informants to "keep talking", whenever they fell silent for longer than 10 seconds. The informants listened to the input text and verbalised their thoughts from the moment on when they started dealing with the respective items that referred to the text they had just listened to, usually by reading the item stem and trying to identify the most plausible option.

This study focuses on two sets of eight multiple-choice items. The first set of items assesses the test-takers' comprehension of eight dialogues between a female and a male speaker. Each item contains three answer options. The second set of items focuses on a narrative that resembles a segment of a radio show. Each item contains four options. Test-takers were allowed to hear all listening texts twice and to correct their item response after having heard the respective text for the second time. This procedure reproduces the specifications used in the DESI main study. Depending on the individual informant's decision whether the first response should be modified or not, this means that for each item a maximum of two responses per informant were recorded in the verbal protocols for each of the 16 items.

Directly after the informants had finished working on the item(s) that referred to one of the listening texts, they were asked to give an oral summary of the text. These oral summaries, which all informants but one chose to verbalise in German, were analysed as evidence of the information the test-takers had integrated into their mental representations of the listening text. These data on the test-takers' comprehension supported the analysis of the think-aloud protocols. In some cases

where informants had stopped verbalising their thoughts for longer stretches of the protocol the comprehension revealed in the oral summaries provided further hints as to how informants may have arrived at choosing the correct (or an incorrect) answer option of the item.

The second data source that was intended to support and confirm the analysis of the think-aloud data drew on stimulated recall interviews that were carried out after the informants had finished working on the two sets of listening items. This interview technique tries to engage informants in retrospective introspection by providing a stimulus that shall activate memory structures concerning previous cognitive processing (cf. Bloom 1954; Di Pardo 1994; Gass & Mackey 2000). In the present study, the stimuli were provided by playing back those recorded segments of the informant's verbal protocol that coincided with the items the informant had identified as the most difficult in the respective set of items. After the think-aloud segment had been played back the informants were asked to speculate why the particular item had posed a problem to them. This intervention deviates from the procedure as laid out by Gass and Mackey in so far as it asks informants to focus on a specific issue. It was decided that the specific focus on test-takers' assessment of item difficulty, which is a research objective that emerged from a phase of piloting think-aloud instructions and possible interview questions, would legitimize this deviation from the *lege artis* procedure. In a final step possible comprehension problems were examined. Informants were handed scripts of the listening text the item referred to in order to allow for a closer look at segments of the text the informant may not have processed successfully while listening.

In summary, the methodological decisions with regard to data elicitation techniques aimed at providing a comprehensive account of the process level of taking the listening test. While the think-aloud data generally help explain why informants were successful in solving a given item, the verbal protocols sometimes lacked information that would explain why informants failed to find the correct answer option. Data from oral summaries and stimulated recall interviews contribute to such explanations.

2.4 Qualitative data analysis

Three different kinds of qualitative data emerged from the phase of data elicitation: Concurrent and immediately retrospective think-aloud protocols, i.e. thoughts verbalised while active in cognition or with a few seconds delay, oral summaries of the listening texts and stimulated recall interview data. The data were transcribed according to the specifications of a reduced version of GAT, a transcription standard commonly used in conversation analysis (cf. Selting et al., 1998). The verbal protocols were then integrated into a database that was then analysed using MAXqda, a software package for qualitative data analysis (cf. Lewins & Silver 2009: 252 ff.)

The main focus of data analysis was on the think-aloud data, which provided the closest link with the cognitive processes elicited by the listening test tasks. As indicated above, oral summaries and stimulated recall interview data were analysed to substantiate the coding of the occasionally patchy think-aloud protocols.

The methodological blueprints for verbal protocol analysis in language testing research offered by Green (1998) were adapted to the context of the present study. Green suggests that verbal data should be segmented on the level of propositions and that the development of a coding scheme should be based on generating code categories inductively, keeping theoretical assumptions to a minimum to allow for the study of task processing that is possibly inconsistent with the approach of the assessment instrument (cf. Green 1998:73). Since the main objective of the present study is to verify the match between task processing and the theoretical assumptions that both task development and the interpretations of test scores are based on, data analysis must be open for processes that are inconsistent with these propositions. On the other hand, codes that describe task processing consistent with the construct must be mapped onto the relevant theoretical terminology in the phase of data analysis. According to Kasper (1998), this theory-based approach to data analysis is especially relevant for the study of cognitive processes:

Because cognitive processes are only indirectly and partially represented in verbal reports, it is necessary to analyze protocols by means of a coding scheme that will guide the researcher's inferences in a principled, theory-based manner.

(Kasper 1998:359)

The central theoretical frame of reference for the study is the research on L2 listening comprehension processes and strategies that informed the development of the DESI EFL listening construct (Buck 1991, 1992, 2001; Kintsch 1998; Ross 1997; Rost 2002). Consequently, the core of the coding scheme reflects relevant components of the models and taxonomies developed in these studies on mostly automated and unconscious language processing, e.g. recalling text information, making inferences, relating perceived information to language knowledge and background knowledge to construct meaning etc. Additionally, phenomena were identified in the verbal protocols that are not defined in the construct. Informants use metacognitive and affective strategies, which they appear to select more or less consciously, to monitor and evaluate their developing comprehension and task performance.

Faced with difficulties in relating their mental representations of the listening texts to the answer options of the test items, informants turn to compensatory test-taking strategies, such as eliminating implausible answer options, using knowledge gained from other items as clues, or selecting an option due to a key word that seems to relate to the listening text (cf. Cohen 1998:103). Coding categories

for such strategic processing behaviour were developed “in vivo”, i.e. in the inductive strand of data analysis, and later linked with concepts that have emerged from research on strategies of language use, language learning and test-taking (Cohen 2000, 2007; O’Malley & Chamot 1990; Vandergrift 2003).

3. Results of qualitative data analysis: Coding verbal reports for cognitive processes and strategy use

The coding scheme contains a total of 156 codes which are grouped according to 12 types of categories that provide information on the informant, the listening task, the item response, the mental operations that preceded the item response and on the test-taker’s retrospective and metacognitive comments on his or her performance on a given item.³ The following section illustrates a selection of those categories of the coding scheme which are central to the investigation of the validity of the listening tasks.

3.1 Central categories of the coding scheme: Recall propositions

The verbal protocols on listening task processing begin right after the informant has finished listening to the text for the first time. Informants generally begin to verbalise their thoughts when they read the stem of the multiple-choice item. The data show that test-takers are faced with the complex task of relating their understanding of the listening text to the information provided in the item. At the beginning of processing the task informants tend to recall information from the listening text and subsequently testing the presumed accuracy of their comprehension and the relevance of the information recalled for the task at hand. In the vast majority of cases it seems that the informant’s mental representation of the listening text, at this early stage, is far from complete and in a rather fragile state of development.

The construction-integration model of text comprehension (Kintsch & van Dijk 1978; Kintsch 1998) that has informed the development of the DESI EFL-listening construct provides the term “proposition” as a feasible unit of analysis both for the meaning-focus of the test items (What does the test-taker have to understand in order to choose the correct response?) and for the facets of meaning the informants verbalize in their think-aloud protocols (What is the nature of the test-taker’s understanding of the listening text? Which facets of his/her understanding of the text does the test-taker deem relevant for his/her attempts at solving the task?).

3. An in-depth description of the coding scheme was published in Rossa 2012 (pp. 115–153).

Propositions are conceptualised as “the semantic processing units of the mind” (Kintsch 1998: 69), encompassing both semantic relations within clauses and functional relations between clauses and text passages (cf. Tirkkonen-Condit 1991: 239). In models of text comprehension, propositions are defined as information that consists of a predicate (e.g. FIND) and one or more arguments (e.g. MAN; THE CD). In the context of this study, a proposition is defined as the verbalisation of recalled information that contains at least a predicate and one argument and consists of up three clauses. This definition includes both “atomic” and “complex” propositions as described by Kintsch (1998: 37–38), but unlike common practice in experimental studies that seek to test models of text comprehension the structure of the propositions, i.e. transcribed think-aloud verbalisations, is kept intact. This makes it possible to analyse the test-takers’ language recall processes, which make up the core of test construct, and their interaction with other mental operations in the wider context of processing a listening task.

A prototypical example of verbal data that were coded as “recall proposition” is present in the following segment of a sample think-aloud protocol transcript. The test-taker reads the item stem in line 299 and recalls a proposition in line 300 that he seems to deem relevant to the accomplishment of the task.

299 ok where did the man find the cd-
300 he had forgotten where his cd is-

The idea that the protagonist of the listening text has forgotten where he might have misplaced a newly-bought compact disc is the first proposition that is presented in this text, introducing the problem the dialogue will focus on. If the test-taker recalls “they said something about a coffee-table”, this utterance is coded as “recall fragment of a proposition”, because one of the propositions in the dialogue presents the information that the woman asks the man to look for his CD on the coffee-table.

From the point of view of cognitive theories of language processing (cf. Anderson 1995; Gernsbacher & Foertsch 1999; Graesser, Gernsbacher & Goldman 1997; Kintsch 1998; Rogers & McClelland 2008) the informants’ verbalisations reflect the results of the perceptual and constructive comprehension processes that were active while they were listening. These propositions and fragments of propositions reflect the information the informants have already integrated into their mental representations of the text.

The informants are continually assessing the truth, relevance and plausibility of these constituents of their comprehension in light of their understanding of the task. The semantic and pragmatic analysis of the information perceived seems to continue into the phase of task processing, as several informants discard

propositions they recall in favour of other, constructed propositions that seem more plausible. Sometimes information in the item seems to prompt such inferential reinterpretations of the text information (cf. following section on coding category “Generate Inferences”).

The coding category that is of central relevance to the validity of the item responses elicited by the tasks refers to instances in which the test-taker recalls the information the item focuses on. Gary Buck defines this as the “necessary information (NI)” (Buck 2001:129), and the codes of this category refer to his concept in their labels: “Recall NI propositions” and “Recall NI fragment”. If the task-processing elicited by the test tasks is a valid reflection of the test construct, then test-takers, who recall the information necessary to answer the item successfully, should be able to select the correct option. If test-takers only recall fragments of the NI, they should not be in a position to select the correct response.

3.2 Central categories of the coding scheme: Generate inferences

The data also show that test-takers continue to process the propositions and fragments they recall, relating them to their background knowledge to construct additional propositions in an inferential process. Research on text comprehension emphasises the central role inferences play in the process of constructing a coherent mental model of the text.

Inferences are generated in three different contexts of task-processing. First, test-takers have to construct coherence in their mental representations of the listening text. These inferential processes are also specified in the DESI listening test construct as the ability to match “explicitly and implicitly presented information (actions, emotions, intentions) with language knowledge and background knowledge to recognize and retrieve, to infer, and to interpret this information” (Nold, Rossa, & Hartig 2008:99). The second context focuses on assessing the possible links between propositions presented in the stem of the item and those suggested in the answer options. These inferences are limited to dealing with the test format, multiple-choice questions. The third context is positioned at an interface between construct-relevant processing and test format processing. Test-takers have to match the answer option of their choice with propositions recalled from the listening text, which requires inferential processing, whenever there is little or no lexical overlap between the correct option and the NI.

Four types of inferences become apparent in the think-aloud protocols: bridging, elaborative, reconstructive and confabulating inferences.

The first type of inferences, bridging inferences, creates coherence by making anaphoric references between propositions explicit (cf. Graesser, Singer, & Trabasso 1994). Bridging inferences are an essential prerequisite for comprehension

on the sentence level of the input text and for basic comprehension of the information provided in the items.

Elaborative inferences are mainly generated to support the coherence of the mental model of the text. In many cases this is obviously motivated by gaps in the test-taker's understanding that he or she is aware of. Test-takers elaborate on what they have gathered from the text and fill gaps in their mental model by constructing propositions based on interpretations of implicit information (cf. Graesser, Wiemer-Hastings, & Wiemer-Hastings 2001; McKoon & Ratcliff 1986). Elaborative inferences are also present in the verbal protocols, whenever test-takers match answer options with relevant propositions recalled from the text.

When test-takers build interpretations based on fragments of explicit information presented in the text, it seems as if they are trying to reconstruct text propositions they were not able to comprehend completely in the listening phase. The ensuing reconstructive inferences tend to be incongruent with the intended meaning of the text, because the textual foundations they are built on are often insufficient for the construction of plausible interpretations.

The coding category “confabulating inferences” describes propositions that rely on predictions the test-takers make. These predictions are often based on the test-takers' overall comprehension of the situation presented in the text, because in most cases these confabulations cannot be traced back directly to information explicitly presented in the text. The term “confabulation” is taken from research in neuropsychology focussing on patients with severe cognitive deficits who spontaneously produce narratives of events which contradict those events that were actually presented in psychological experiments. Roser and Gazzaniga conclude that these “bizarre deficits of consciousness [...] probably result from interpretations of incomplete information” (2004: 57). In contrast to the phenomena identified in the narratives produced by brain lesion patients, confabulating inferences discovered in the verbal data of the test-takers show varying degrees of plausibility from the point of view of the intended meanings of the text. They often point to invented phenomena outside the original input text which seem to support the coherence of the situation model, e.g. predicting what may happen after the dialogue or narrative presented in the listening text ends. These inferences emphasise the creative nature of the cognitive processes they result from, but the analysis of the verbal protocols shows that they are, in fact, occasionally compatible with the intended meaning of the text (16 instances out of a total of 70 codings: 22,88 %). Confabulations in line with the intended meanings of the input text tend to coincide with successful item responses. Only in one case the test-taker cannot make use of his confabulating inference, because she fails to decode one of the distractors correctly and misguidedly matches it with her inference.

The following sample transcript illustrates two core categories of task processing: making an elaborative inference, and recalling the necessary information (NI). The task refers to a narrative text about a character named Mr Vialli, who boards a plane in San Francisco to fly home to Rome, Italy. His plane stops in New York to refuel and Vialli gets out, thinking he had arrived in Rome. He walks around the city and wonders why Rome's historic sights have disappeared. The test-taker, who picked "cent" as his alias, works on an item that targets the comprehension of the main idea that Vialli is not in Rome but rather, as the correct answer option puts it, "in a different city".

Cent starts reading aloud the stem (line 383 of the transcript) and the answer options of the multiple-choice item (lines 384–387). It seems that choice number two, "the buildings were no longer there", presents a reading comprehension problem for the informant. Cent hesitates while reading the option and decides to ignore the word "no".

383 cent: mister vialli couldn't find
 rome's famous buildings because-
 384 the way to the buildings was too long-
 385 the buildings were errm longer there-
 386 he was in a different city-
 387 he didn't have a map of city.

Cent opts for "he was in a different city" as the correct answer choice in line 388 and tries to support the legitimacy of his response by assessing the plausibility of the other distracters.

388 cent: ok he was in a different city.
 389 int: hmhm-
 390 cent: it makes no different-
 391 everywhere gibt-
 there are difference buildings.
 392 int: hmhm-
 393 cent: but we don't know
 ok i think in italian can be buildings-
 394 that longer as americans buildings.
 395 int: hmhm-
 396 cent: as new york's but-

This strategy prompts an elaborative inference. In an instance of English-German code mixing the informant asserts in lines 391 ff. that choice number two, "the

buildings were no longer there”, cannot be true according to his background knowledge, because different kinds of buildings can be found (German: *gibt es*) “everywhere”. His elaborative inference states that buildings in Italy might be just as “long” as those in New York City.

Cent was obviously unable to decode the meaning of the words “no longer” correctly, as indicated by the hesitation phenomenon pointed out above. Cent generates an inference the distractor was not meant to elicit, provoked by a decoding problem in reading the information presented in the item. Luckily, Cent’s reading comprehension problem and the elaborative inference he generates do not motivate him to re-assess his answer choice. Apparently, Cent evaluates his inference as evidence against the plausibility of distractor number two. Ultimately, the informant recalls the necessary information (cf. line 399) and is able to match his understanding of the main idea with the correct option “he was in a different city”.

397 cent: he was in a different city.

398 int: hmhm-

399 cent: he wasn't in the same city that he wanted-

From the point of view of validation research this episode can be interpreted in favour of the validity of this item. Despite the potentially disturbing effect of the reading comprehension problem caused by the wording of distractor number two the construct-relevant process of recalling the NI has led to a successful item response in this case. Earlier in the same transcript the informant demonstrates that his comprehension of the main idea as assessed in this item is additionally substantiated by an elaborative inference (line 381) that creates coherence on the global level of the text:

380 cent: okay he thought he was in italy.

381 because he was a crazy man.

A total of 329 item responses were coded in the verbal protocols as either “no response”, “select incorrect option” or “select correct option”. The verbalisations that preceded each item response were coded for the mental operations pertaining to task processing.

In five cases, test-takers selected an incorrect answer option after they had listened to the text for the second time, although they had selected the correct option before. Interestingly, the four informants that demonstrated this phenomenon all belonged to the subsample whose general foreign language ability was particularly weak.

3.3 Central categories of the coding scheme: Employ test-taking strategies

The qualitative data analysis reveals a considerable number of mental operations test-takers are engaged in during task processing which do not reflect the procedural facets of the test construct specified for the listening test. Most evidently, test-takers respond to the demands of the test format by activating test-taking strategies, which Cohen defines as “those test-taking processes that the respondents have selected and of which they are conscious, at least to some degree” (Cohen 1998: 92).

In the present study the term test-taking strategies is confined to those strategies that test-takers use to “opt out of the language task at hand” or “circumvent the need to tap their actual language knowledge or lack of it” (ibid.). Cohen’s research on test-taking strategies (cf. Cohen 2000, 2007) provides categories for such construct-irrelevant operations that may negatively influence the measurement outcome, if these processes mask test-takers’ true abilities (cf. Haladyna & Downing 2004; Van der Veen, Huff, Gierl, McNamara, Louwerse, & Graesser 2007: 140).

When test-takers are faced with multiple-choice items that seem too difficult, the most obvious test-taking strategy that allows them to opt out of the task is guessing which answer option is correct. The data contain 28 instances of guessing which do not provide any evidence for the reasons why test-takers may have favoured one answer option over the other. Generally, the informants comment on their response, arguing that they were forced to guess, either because they do not know the specific information the item asks for, or because they do not understand the question.

In 34 cases the verbal data show that test-takers’ guesses are based on fragments of the original text that they are able to recall and match with words presented in the answer options. These responses are coded as “matching fragment and option”.

The verbal protocols also contain phenomena that indicate a reversal of the process of guessing. Test-takers speculate about possible text information they may have missed while listening and base these speculations on the information presented in the item. Test-takers from both subsamples seem to be involved in matching their understanding of the text with their understanding of the item, but informants from subsample LO show a much stronger tendency to trust information deduced from item stems and answer options more than their own, often unstable mental model of the listening text.

A rough analysis of the construct-irrelevant test-taking strategies illustrated above provides arguments for the general validity of the listening tasks. Guessing, for example, coincides with selecting the correct answer option in only three out of 28 cases. Matching a fragment recalled from the text with an answer option results in a successful completion of the item in nine out of 34 cases.

4. Discussion of research results

The analysis of the verbal protocols produces a complex image of the mental operations test-takers engage in while they attempt to accomplish the listening test tasks. With regard to research questions one and two, which focus on the extent to which the mental operations elicited by the listening test tasks correspond to the specifications of the test construct, the data show that there is much more to task-processing than what the test seeks to measure. It is obvious that a considerable part of each test-taker's processing capacity is used on the task of understanding what the multiple-choice items may be asking for and selecting the most plausible answer option.

Even within the constraints of the multiple-choice format, a range of mental operations can precede the successful selection of the correct answer option. For some test-takers this may mean going back to almost all the propositions they can recall from the original text before they can decide on an answer option. Others may have to add elaborative inferences to their patchy mental representation of the presumed idea the item focuses on, before they can respond to the item. Those test-takers that have developed a high level of general proficiency in the L2, on the other hand, tend to recall the necessary information, and they can generally match their own understanding of the targeted information with the propositions provided in the correct answer option with ease. These findings support the notion that it is highly problematic to specify the subskills, such as generating inferences, and strategic processing that are supposedly measured by individual items (cf. Brindley 1998: 127).

The validity of one particular item is called into question by the fact that two test-takers from subsample HI, who had obviously recalled most of the necessary information, nevertheless ruled out the correct option, arguing that the actions implied in this option would go against what would normally happen in such situations, according to their experiential knowledge.

The third research question aims at a general evaluation of the validity of the listening tasks in the context of the probabilistic measurement model that was applied to the test: What is the nature of test-takers' mental processing that coincides with correct item responses? The qualitative analysis of the verbal data yields positive findings with regard to the contribution of construct-relevant processes to successful item responses. In order to examine the dependability of these findings, the statistical relationships between codings of mental operations during task processing and item responses were subjected to Pearson's chi-square tests of independence (cf. Pearson 1900).

This procedure tests, whether paired observations on two variables are independent of each other. The chi-square test assumes total independence of the two

variables as the null hypothesis, the alternative hypothesis states that a relationship between the two variables does exist. Conventionally, the critical p-value which implies that the null hypothesis should be rejected is defined at 5%. For those variables that appear to be related Yule's phi-coefficient is calculated to provide a measure of the relative strength of these relationships. The following tables report the observed frequencies of the central categories of the coding scheme illustrated above, followed by the computation of the respective chi-square and phi values. The variables relating to task processing were recoded to meet the requirement of the chi square test that all expected cell frequencies should be equal to or greater than five. In the process of recoding the data, all instances of the specified mental operation (e.g. generating one or more than one inference) preceding an item response were subsumed in the value one. Additionally, both selecting an incorrect answer option and not selecting any option were subsumed in the value zero for the variable "item response: correct option".

Table 2 focuses on the relationship between recalling the necessary information, the most relevant facet of the listening construct, and task accomplishment.

Table 2. Concurrence of recalling the NI and successful item responses

Descriptive statistics: contingency table			Item response: correct option		Total
			0	1	
Cognitive processing: recall NI propositions [dichotomised]	0	N	97	54	151
		%	81.5%*	25.7%	45.9%
	1	N	22	156	178
		%	18.5%	74.3%	54.1%
Total		N	119	210	329
		%	100.0%	100.0%	100.0%

*percentages refer to columns

Measures of association:

Chi-Square and Phi coefficients

Statistics	Value	df	Significance	
			Asymptotic ^b	Approximated
Pearson Chi-Square	95.237a	1	.000	
Yule Phi	.538			.000
Valid cases	329			

a. No cells have expected count less than 5. The minimum expected count is 54.62.

b. 2-sided

In nearly 75% of the cases, successful item responses follow from the test-takers' recall of at least one complete proposition related to the necessary information. As indicated by the phi value of .538, this relationship is moderately strong across all 16 items from the perspective of the 329 item responses recorded. This finding suggests that success on the listening test does depend largely on the ability to understand and recall relevant information from the listening text, as implied in the test construct.

Table 3 reports the frequencies of inferential comprehension processes in successful and failed task processing. The descriptive statistics suggest a moderate relationship between the execution of inferential comprehension processes and task accomplishment. Combining text information and world knowledge is not as strongly connected to correct item responses as the recall of the necessary information.

These findings substantiate the results of qualitative data analysis that describe a striking variability among informants and between items with regard to the generation and elicitation of inferences: Some test-takers accomplish test tasks without feeling the explicit need to engage in inferential comprehension processes,

Table 3. Concurrence of inferential processing and successful item responses

Descriptive statistics: contingency table			Item response: correct option		Total
			0	1	
Cognitive processing: recall NI propositions [dichotomised]	0	N	82	73	155
		%	68.9%	34.8%	47.1%
	1	N	37	137	174
		%	31.1%	65.2%	52.9%
Total		N	119	210	329
		%	100.0%	100.0%	100.0%

Measures of association:

Chi-Square and Phi coefficients

Statistics	Value	df	Significance	
			Asymptotic ^b	Approximated
Pearson Chi-Square	35.543 ^a	1	.000	
Yule Phi	.329			.000
Valid cases	329			

a. No cells have expected count less than 5. The minimum expected count is 56.06.

b. 2-sided

and some items do not require the test-taker to draw any more than the most elementary bridging inferences. The following tables report descriptive statistics for these observations.

Table 4a shows that informants who have acquired a high level of general foreign language ability tend to generate more inferences than informants from subsample LO. The relatively high variance in subsample LO with regard to the number of inferences the informants make reflects the observation that these test-takers show a high degree of individual differences regarding strategy use. Informants at the lower end of the ability scale seem to experience some tasks as excessively demanding. They report complete breakdowns of comprehension and seem to have no text information available they could draw inferences from. Other informants from subsample LO, however, seem to be well-versed in the use of inferences as compensatory strategies to make up for gaps in their understanding of the text and generate a lot of inferences, hoping that some of them may point them towards the correct answer option.

Table 4a. Descriptive statistics for the occurrence of inferences with regard to the subsamples

	Max	Min	M	SD	Variance
Subsample HI	34	15	19,78	5,97	35,69
Subsample LO	24	1	12,11	7,52	56,61
Total			15,94	7,68	59,00

Table 4b provides information on differences among the two task types in the test: items that focus on dialogues and items that refer to a longer narrative. The relatively high number of inferences elicited by the dialogue items can probably be explained by the missing context that the listener has to construct for each of these very short dialogues.

Table 4b. Descriptive statistics for the occurrence of inferences with regard to the two types of listening items used in the test

	Max	Min	M	SD	Variance
Dialogue items	27	14	20,75	4,65	21,64
Narrative items	24	7	15,13	4,76	22,70
Total			17,94	5,40	29,13

The most obvious threat for the validity of multiple-choice items lies in the successful application of a test-taking strategy that conceals the actual language

knowledge of the test-takers: guessing the correct answer. The descriptive statistics for instances of guessing that yield successful item responses clearly demonstrate that this threat can be considered insignificant for the DESI listening items (see Table 5). This interpretation is supported by the moderately negative value of the phi coefficient.

Table 5. Concurrence of guessing the correct answer and successful item responses

Descriptive statistics: contingency table			Item response: correct option		Total
			0	1	
Cognitive processing: recall NI propositions [dichotomised]	0	N	94	207	301
		%	79.0%	98.6%	91.5%
	1	N	25	3	28
		%	21.0%	1.4%	8.5%
Total		N	119	210	329
		%	100.0%	100.0%	100.0%

Measures of association: Chi-Square and Phi coefficients			Significance	
Statistics	Value	df	Asymptotic ^b	Approximated
Pearson Chi-Square	37.399 ^a	1	.000	
Yule Phi	-.337			.000
Valid cases	329			

a. No cells have expected count less than 5. The minimum expected count is 10.13.

b. 2-sided

5. Conclusions and Implications

This study has made use of a radically reduced concept of validity to investigate the question how performance on a sample of test items mirrors the facets of L2 listening ability specified in the test construct. Verbal data on test-takers' task processing were collected from two extreme subsamples of particularly high and low scores on a C-test. Data were coded with regard to the theoretical frame of reference that had guided task development. Additional categories were developed inductively in the process of data analysis and linked to concepts that were previously identified in relevant research studies on test-taking processes and strategies of language use.

The qualitative analysis of the verbal data shows that the multiple-choice test format requires processing that deviates from the processes predicted in the test construct. This finding is in alignment with an exploratory study into the underlying construct of a reading test. Rupp et al. call attention to the fact “that different MC questions do not merely tap but, indeed, create very particular comprehension and response processes” (Rupp, Ferne, & Choi 2006: 470). Three items seem to elicit response processes that constitute an obvious threat to validity:

- recalling the necessary information, but not selecting the correct option,
- recalling information that contradicts the meaning of the necessary information, but still selecting the correct option.

These items should obviously be revised or discarded from the test. But despite the complexity and variability of the response processes as explored in the present study, two core facets of the listening construct – recalling information from the listening text and generating appropriate inferences – do play an important role in determining the measurement outcome, while test-taking strategies, such as guessing, do not. This result can reasonably be interpreted as an argument to support the general validity of the DESI listening test tasks.

The quantitative analysis of the coded segments in the verbal protocols indicates that the think-aloud procedure has effectively allowed insights into informants’ cognitive processing during task performance. Among a total of 210 episodes of think-aloud data that end with the test taker choosing the correct option, for example, only nine contain no information on propositions the test taker may have recalled from listening to the text and deemed relevant for the process of selecting the correct option.

The study is obviously limited with regard to the number of test items investigated and is based on a small sample size due to the qualitative approach employed. Nevertheless, further research could build on the research design employed in the present study and investigate the task processing elicited by other test formats. Compiling a body of evidence on the attributes that actually determine measurement outcomes of specific tasks can inform the development of theoretical models of test task processing to account for variations among task characteristics and test-takers’ abilities.

References

- Anderson, J. R. (1995). *Cognitive psychology and its implications* (4th ed.). New York, NY: W.H. Freeman and Company.
- Bachman, L. F. (2004). *Statistical analyses for language testing*. Cambridge: Cambridge University Press.

- Bachman, L. F. (2005). Building and supporting a case for test use. *Language Assessment Quarterly*, 2, 1–34. doi:10.1207/s15434311laq0201_1
- Bloom, B. (1954). The thought processes of students in discussion. In S. J. French (Ed.), *Accent on teaching: Experiments in general education* (23–46). New York, NY: Harper.
- Blumer, H. (1954). What is wrong with social theory? *American Sociological Review*, 18, 3–10. doi:10.2307/2088165
- Borsboom, D., Cramer, A. O. J., Kievit, R. A., Zand Scholten, A., & Franic, S. (2009). The end of construct validity. In R. W. Lissitz (Ed.), *The concept of validity: Revisions, new directions, and applications* (pp. 135–170). Charlotte, NC: Information Age Publishers.
- Borsboom, D., VanHeerden, J., & Mellenbergh, G. (2004). The concept of validity. *Psychological Review*, 111(4), 1061–1071. doi:10.1037/0033-295X.111.4.1061
- Brindley, G. (1998). Describing language development? Rating scales and SLA. In L. F. Bachman & A. D. Cohen (Eds.), *Interfaces between second language acquisition and language testing research* (pp. 112–141). Cambridge: Cambridge University Press.
- Bryman, A. (2006). Integrating quantitative and qualitative research: How is it done? *Qualitative Research*, 6(1), 97–113. doi:10.1177/1468794106058877
- Buck, G. (1991). The testing of listening comprehension: An introspective study. *Language Testing*, 8(1), 67–91. doi:10.1177/026553229100800105
- Buck, G. (1992). Listening comprehension: Construct validity and trait characteristics. *Language Learning*, 42(3), 313–357. doi:10.1111/j.1467-1770.1992.tb01339.x
- Buck, G. (2001). *Assessing listening*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511732959
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multi-trait-multimethod matrix. *Psychological Bulletin*, 56, 81–105. doi:10.1037/h0046016
- Chapelle, C. A. (1998). Construct definition and validity inquiry in SLA and research. In L. F. Bachman & A. D. Cohen (Eds.), *Interfaces between second language acquisition and language testing research* (pp. 32–70). Cambridge: Cambridge University Press.
- Cohen, A. D. (1998). Strategies and processes in test taking and SLA. In L. F. Bachman & A. D. Cohen (Eds.) *Interfaces between second language acquisition and language testing research* (pp. 90–111). Cambridge: Cambridge University Press.
- Cohen, A. D. (2000). Exploring strategies in test-taking: Fine-tuning verbal reports from respondents. In G. Ekbatani & H. Pierson (Eds.), *Learner-directed assessment in ESL* (pp. 127–150). Mahwah, NJ: Lawrence Erlbaum Associates.
- Cohen, A. D. (2007). The coming of age for research on test-taking strategies. In J. Fox, M. Weshe, D. Bayliss, L. Cheng, C. Turner, & C. Doe (Eds.), *Language testing reconsidered* (pp. 80–111). Ottawa: Ottawa University Press.
- Creswell, J. W., Plano Clark, V. L., Gutmann, M. L., & Hanson, W. E. (2007). Advanced mixed methods research designs. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of mixed methods in social & behavioral research* (pp. 209–240). Thousand Oaks, CA: Sage.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin* 52(1), 281–302. doi:10.1037/h0040957
- Denzin, N. K. (1970). *The research act: A theoretical introduction to sociological methods*. Englewood Cliffs, NJ: Prentice Hall.
- Denzin, N. K., & Lincoln, Y. S. (Eds.). (2000). *The handbook of qualitative research* (2nd ed.). Thousand Oaks, CA: Sage.
- DESI-Konsortium (Ed.). (2008). *Unterricht und Kompetenzerwerb in Deutsch und Englisch: Ergebnisse der DESI-Studie*. Weinheim: Beltz.

- Di Pardo, A. (1994). Stimulated recall in research on writing: An antidote to "I don't know, it was fine". In P. Smagorinsky (Ed.), *Speaking about writing: Reflections on research methodology* (pp. 163–184). Thousand Oaks, CA: Sage.
- Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data* (Rev. ed.). Cambridge, MA: The MIT Press.
- Ericsson, K. A. (2003). Valid and non-reactive verbalisation of thoughts during performance of tasks: Toward a solution to the central problems of introspection as a source of scientific data. *Journal of Consciousness Studies*, 10(9–10), 1–18.
- Friese, M., & Fiedler, K. (2010). Being on the lookout for validity. *Experimental Psychology*, 57(3), 228–232. doi:10.1027/1618-3169/a000051
- Gass, S. M., & Mackey, A. (2000). *Stimulated recall methodology in second language research*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Gernsbacher, M. A., & Foertsch, J. A. (1999). Three models of discourse comprehension. In S. Garrod, & M. J. Pickering (Eds.), *Language processing* (pp. 283–299). Hove: Psychology Press.
- Graesser, A. C., Gernsbacher, M. A., & Goldman, S. R. (1997). Cognition. In T. A. van Dijk (Ed.), *Discourse studies. A multidisciplinary introduction*. Vol.1 (pp. 292–319). Thousand Oaks, CA: Sage. doi:10.4135/9781446221884.n11
- Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, 101, 371–95. doi:10.1037/0033-295X.101.3.371
- Graesser, A. C., Wiemer-Hastings, P., & Wiemer-Hastings, K. (2001). Constructing inferences and relations during text comprehension. In T. Sanders, J. Schilperoord, & W. Spooren (Eds.), *Text representation: Linguistic and psycholinguistic aspects* (pp. 249–271). Amsterdam: John Benjamins. doi:10.1075/hcp.8.14gra
- Green, A. (1998). *Verbal protocol analysis in language testing research: A handbook*. Cambridge: Cambridge University Press.
- Grotjahn, R., & Eckes, T. (2006). A closer look at the construct validity of C-tests. *Language Testing*, 23(3), 290–325. doi:10.1191/0265532206lt3300a
- Haastруп, K. (1987). Using thinking aloud and retrospection to uncover learners' lexical inferring procedures. In C. Faerch & G. Kasper (Eds.), *Introspection in second language research* (pp. 197–212). Clevedon: Multilingual Matters.
- Haladyna, T. M., & Downing, S. M. (2004). Construct-irrelevant variance in high-stakes testing. *Educational Measurement: Issues and Practice*, 23(1), 17–27. doi:10.1111/j.1745-3992.2004.tb00149.x
- Kane, M. (2001). Current concerns in validity theory. *Journal of Educational Measurement*, 38(4), 319–342. doi:10.1111/j.1745-3984.2001.tb01130.x
- Kasper, G. (1998). Analysing verbal protocols. *TESOL Quarterly*, 32(2), 358–362.
- Kintsch, W. (1998). *Comprehension. A paradigm for cognition*. Cambridge: Cambridge University Press.
- Kintsch, W., Patel, V. L., & Ericsson, K. A. (1999). The role of long-term working memory in text comprehension. *Psychologia*, 42, 186–198.
- Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review*, 85, 363–394. doi:10.1037/0033-295X.85.5.363
- Kunnan, A. J. (2000). Fairness and justice for all. In A. J. Kunnan (Ed.), *Fairness and validation in language assessment* (pp. 1–14). Cambridge: Cambridge University Press.
- Lazarsfeld, P. F. (1960). *Latent structure analysis and test theory*. In H. Gulliksen & S. Messick (Eds.), *Psychological scaling: Theory and applications* (pp. 83–96). New York, NY: Wiley.

- Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: The MIT Press.
- Lewins, A., & Silver, C. (2009). *Using software in qualitative research: A step-by-step guide* (Reprinted.). Los Angeles: Sage.
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Maxwell, J. A. (2005). *Qualitative research design: An interactive approach*. Thousand Oaks, CA: Sage.
- McKoon, G., & Ratcliff, R. (1986). Inferences about predictable events. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 12, 82–91. doi:10.1037/0278-7393.12.1.82
- McNamara, T. F., & Roever, R. (2006). *Language testing. The social dimension*. Malden, MA: Blackwell.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed.; pp. 13–103). New York, NY: American Council on Education & Macmillan.
- Messick, S. (1992). Validity of test interpretation and use. In M. C. Alkin (Ed.), *Encyclopedia of educational research* (pp. 88–98). New York, NY: Macmillan.
- Messick, S. (1996). Validity and washback in language testing. *Language Testing*, 13(3), 241–256. doi:10.1177/026553229601300302
- Miles, M. B., & Huberman, A. M. (2009). *Qualitative data analysis: An expanded sourcebook*. Thousand Oaks, CA: Sage.
- Mislevy, R. J. (1996). Test theory reconceived. *Journal of Educational Measurement*, 33(4), 379–416. doi:10.1111/j.1745-3984.1996.tb00498.x
- Nold, G., & Rosssa, H. (2007). Hörverstehen. In B. Beck & E. Klieme (Eds.), *Sprachliche Kompetenzen. Konzepte und Messung – DESI-Studie (Deutsch-Englisch-Schülerleistungen International)* (pp. 178–196). Weinheim: Beltz.
- Nold, G., Rossa, H., & Hartig, J. (2008). Proficiency scaling in DESI listening and reading EFL tests: Task characteristics, item difficulty and cut-off points. In L. Taylor & C. J. Weir (Eds.), *Multilingualism and assessment. Achieving transparency, assuring quality, sustaining diversity. Proceedings of the ALTE Berlin conference, May 2005*. (pp. 94–116). Cambridge: Cambridge University Press.
- O'Malley, M., & Chamot, A. U. (1990). *Learning strategies in second language acquisition*. Cambridge: Cambridge University Press. doi:10.1017/CBO9781139524490
- Patton, M. Q. (2002). *Qualitative evaluation and research methods*. Thousand Oaks, CA: Sage.
- Pearson, K. (1900). On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can reasonably be supposed to have arisen from random sampling. *Philosophical Magazine*, 5(50), 157–175. doi:10.1080/14786440009463897
- Pienemann, M. (1998). *Language processing and second language development: Processability Theory*. Amsterdam: John Benjamins. doi:10.1075/sibil.15
- Pienemann, M. (2003). Language processing capacity. In C. Doughty & M. Long (Eds.), *The handbook of second language acquisition* (pp. 679–714). Oxford: Blackwell. doi:10.1002/9780470756492.ch20
- Pienemann, M., & Kessler, J.-U. (2007). Measuring bilingualism. In P. Auer & L. Wei (Eds.), *Handbooks of applied linguistics: Handbook of multilingualism and multilingual communication* (pp. 247–278). Berlin: Mouton de Gruyter.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Danmarks pædagogiske Institut.

- Rasch, G. (1961). *On general laws and the meaning of measurement in psychology*. Berkeley, CA: University of California Press.
- Rasch, G. (1968). *An individualistic approach to item analysis*. In P. F. Lazarsfeld & N. W. Henry (Eds.), *Readings in mathematical social science* (pp. 89–107). Cambridge, MA: The MIT Press.
- Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests* (exp. ed.). Chicago, IL: University of Chicago Press.
- Richards, K. (2003). *Qualitative inquiry in TESOL*. Houndmills: Palgrave.
doi:10.1057/9780230505056
- Rogers, T. T., & McClelland, J. L. (2008). Précis of semantic cognition: A parallel distributed processing approach. *Behavioral and Brain Sciences*, 31(06), 689–714.
doi:10.1017/S0140525X0800589X
- Roser, M., & Gazzaniga, M. S. (2004). Automatic brains – Interpretive minds. *Current Directions in Psychological Science*, 13(2), 56–59. doi:10.1111/j.0963-7214.2004.00274.x
- Ross, S. (1997). An introspective analysis of listener inferencing on a second language listening test. In G. Kasper & E. Kellerman (Eds.), *Communication strategies: Psycholinguistic and sociolinguistic perspectives* (pp. 216–237). Harlow: Addison Wesley Longman.
- Rossa, H. (2012). *Mentale Prozesse beim Hörverstehen in der Fremdsprache. Eine Studie zur Validität der Messung sprachlicher Kompetenzen* (Inquiries in Language Learning, Volume 5). Frankfurt: Peter Lang.
- Rost, M. (2002). *Teaching and researching listening*. Harlow: Pearson Education.
- Rupp, A., Ferne, T., & Choi, H. (2006). How assessing reading comprehension with multiple-choice questions shapes the construct: A cognitive processing perspective. *Language Testing*, 23(4), 441–474. doi:10.1191/0265532206lt3370a
- Senécal, A. (2011). *Processing the L2 comprehension process: Testing Processability Theory's predictions in an ERP study of adult learners of L2 Swedish*. Master's thesis, Lund University. Accessed from: (<http://lup.lub.lu.se/luur/download?func=downloadFile&recordOid=2439371&fileOid=2439378>)
- Selting, M., Auer, P., Barden, B., Bergmann, J. R., Couper-Kuhlen, E., & Günthner, S. et al. (1998). Gesprächsanalytisches Transkriptionssystem (GAT). *Linguistische Berichte*, 173, 91–122.
- Shohamy, E. (2001). *The power of tests: A critical perspective on the uses of language tests*. Harlow: Pearson.
- Smith, J. A. (Ed.). (2003). *Qualitative psychology: A practical guide to research methods*. London: Sage.
- Stoyhoff, S. (2009). Recent developments in language assessment and the case of four large-scale tests of ESOL ability. *Language Teaching*, 42(1), 1–40. doi:10.1017/S0261444808005399
- Tirkkonen-Condit, S. (1991). Relational propositions in text comprehension processes. In K. Sajavaara (Ed.), *Communication and discourse across cultures and languages*. In *AFimLA Yearbook 1990*. (pp. 239–246). Jyväskylä: University of Jyväskylä.
- Van der Veen, A., Huff, K., Gierl, M., McNamara, D. D., Louwerse, M., & Graesser, A. (2007). Developing and validating instructionally relevant reading competency profiles measured by the critical reading section of the SAT reasoning test. In D. S. McNamara (Ed.), *Reading comprehension strategies. Theories, interventions, and technologies* (pp. 137–172). New York, NY: Lawrence Erlbaum Associates.
- Van Someren, M. W., Barnard, Y. F., & Sandberg, J. A. (1994). *The think aloud method: A practical guide to modelling cognitive processes*. London: Academic Press.

- Vandergrift, L. (2003). Orchestrating strategy use: Toward a model of the skilled second language listener. *Language Learning*, 53(3), 463–496. doi:10.1111/1467-9922.00232
- Webb, E., Campbell, D. T., Schwartz, R. D., & Sechrest, L. (1962). *Unobtrusive measures: Nonreactive measures in the social sciences*. Chicago, IL: Rand McNally.
- Weir, C. J. (2005). *Language testing and validation. An evidence-based approach*. Houndmills: Palgrave.
- Yin, R. K. (2003). *Case study research: Design and methods* (3rd ed.). Thousand Oaks, CA: Sage.

About the Authors

Stefanie Couve de Murville is a researcher at the University of Hildesheim. Her involvement within the field of bilingual education has included publishing articles, giving presentations and leading workshops. Her Ph.D., which has a primary school focus, investigates student-teacher interaction in immersion classrooms with the help of conversation analysis.

Email: couves@uni-hildesheim.de

Katharina Hagenfeld works as a research associate at Paderborn University. Her doctoral research focuses on the relationship between the descriptive levels of the Common European Framework of Reference and the developmental stages as explained by Processability Theory.

Email: katharina.hagenfeld@uni-paderborn.de

Dr. Gisela Håkansson is Professor in General Linguistics at Lund University. Her main research interest is language acquisition, and she has led several research projects on the acquisition of Swedish as first and second language and in children with language impairment. She takes a particular interest in language typology and she is the Swedish representative in three EU-funded projects on bilingual children with different language combinations.

Email: Gisela.Hakansson@ling.lu.se

Satomi Kawaguchi, Ph.D., is senior lecturer at the Western Sydney University, and teaches second language acquisition, TESOL research methodology and Japanese as a second language. She has contributed to the recent development in Processability Theory. She is also interested in theory and research-driven-practice, especially the use of new communication technologies and student-focused innovative pedagogies.

Email: S.Kawaguchi@uws.edu.au

Kristin Kersten is professor of EFL and Language Acquisition at the University of Hildesheim. She coordinated the multilateral EU COMENIUS project ELIAS, an international study of bilingual preschools. She has published books and articles on second language acquisition and young learners, language learning, intercultural competence and bilingual environmental education in bilingual preschools,

language input in bilingual classrooms and immersion programs. Her current research in psycholinguistics and EFL focuses on variables of language acquisition, with a special emphasis on bilingual education (CLIL / immersion). She is involved in teacher training at university level and for practitioners in preschools and schools.

Email: kristin.kersten@uni-hildesheim.de

Jörg-U. Kessler Ph.D., is professor of Applied Linguistics and Vice Rector for Academic and International affairs at Ludwigsburg University of Education. His main research interests encompass (instructed) L2 acquisition, linguistic profiling, the application of Processability Theory to the L2 classroom and bilingual language (Immersion & CLIL).

Email: kessler@ph-ludwigsburg.de

Anke Lenzing is assistant professor of English Linguistics at Paderborn University. Her main research interests within SLA are early L2 acquisition and the L2 initial state, L2 transfer, and the interface between comprehension and production.

Email: alenzing@mail.uni-paderborn.de

Mathias Liebner is a teacher for English and Mathematics at the Carl-von-Ossietzky Gymnasium in Bonn, Germany. His research focuses on interlanguage variation, language learner profiling, Task-based Language Teaching and the implementation of Processability Theory in EFL classroom settings.

Email: liebner@me.com

Bo Liu holds a Bachelor's degree on English linguistics and literature and a Master's degree on TESOL (Teach English to Speakers of Other Languages). His teaching areas and research interests are Chinese/English translation and L2 English and Chinese acquisition. He is currently working on her Ph.D. project at the Australian National University. The aim of the project is to investigate the L2 acquisition of Chinese syntax from a processing perspective.

Email: scottliubo@gmail.com

Esther Maier is a primary school teacher. She studied German and English as her main subjects at the University of Hildesheim. During her time as a research assistant at Hildesheim University, she was involved in data elicitation and analysis within the context of the project *Foreign Language Acquisition in Primary School*. In her current post she teaches German as a Second Language for immigrant children.

Email: esther.maier@yahoo.de

Lea Neubauer studied German and English at the University of Hildesheim. Both her Bachelor and Master theses focused on EFL. Within this context she was involved in data elicitation and analysis for the project *Foreign Language Acquisition in Primary School* at Hildesheim University. She now works as a primary school teacher.

Email: glueckliche_lea@gmx.de

Katharina Ponto is a researcher in the field of Language Acquisition and EFL at the University of Hildesheim. Her research interest focuses on language input in primary schools. She coordinates the Project *SMILE* (Studies on Multilingualism In Language Education), a longitudinal study investigating variables of language acquisition in different primary school settings.

Email: schwirz@uni-hildesheim.de

Jana Roos, Ph.D., is Professor of Applied Linguistics and English Language Education at Paderborn University, Germany. Her research focuses on Second Language Acquisition and (Early) Foreign Language Teaching and Learning in Institutional Contexts.

Email: jroos@mail.uni-paderborn.de

Henning Rossa holds a Ph.D. in Applied Linguistics/TEFL and is currently a substitute professor of Applied Linguistics/Foreign Language Education at TU Dortmund University. Henning's research interests include the process level of learners' task performances, early bilingual education and research on language teachers' beliefs.

Email: henning.rossa@tu-dortmund.de

Yanyin Zhang (Ph.D.) is a Senior Lecturer in the Department of East Asian Studies, College of Asia and the Pacific, the Australian National University. She has taught English and Chinese languages, as well as postgraduate TESOL/Applied Linguistics courses at tertiary institutions in Australia, China, Vietnam, and the US. Her current research focuses on Second Language Acquisition and Pedagogy, and international education.

Email: yanyin.zhang@anu.edu.au

Index

- A**
argument mapping 35–36, 39,
55–56, 58
argument structure 3–4, 7–31,
37, 38, 140
a-structure *see* argument
structure
autoprofiling 136, 140, 143–145,
148, 150, 154, 157
- B**
Bad Choice Hypothesis 105
- C**
ceiling effect 185–186, 189
cognitive processes 207, 213,
216–217, 219–220, 223
Common European
Framework of Reference
xvii, 124, 135–160
CEFR *see* Common European
Framework of Reference
communicative tasks xvi, xvii,
14–15, 93, 121, 125, 127, 131, 137,
142, 148, 163–164, 168–170, 172,
182–183, 186, 189, 195
compound tense 70–71
- D**
developmental readiness
122–123, 125, 127, 132
Developmentally-moderated
Transfer Hypothesis xvii,
79, 80
DMTH *see* developmentally
moderated transfer
hypothesis
diagnostic approach xvii,
193, 196
distributional analysis 15, 17,
19–20, 22, 26, 28–29, 46, 86,
88–89, 95, 128, 142
- E**
EFL classroom 121, 127,
193–198, 203, 240
- emergence criterion xi, 69,
83–84, 102, 116, 135–136,
142–143, 145, 156, 182, 185, 202
end-state 101–103, 105, 114, 117
- F**
feasibility 150, 155, 158
Focus on form xi, xvi, xvii,
121–122, 124–128, 130–131, 157,
193, 196–197, 202–203
Focus on formS 113, 117
foreign language
instruction 121
formulae 15, 19, 22, 26, 29, 88,
90–91, 97, 141, 153
formulaic sequences 15, 18, 20,
22–23, 26, 30–31, 152
full transfer 80, 85, 94, 97
- G**
grammatical functions xi,
4–6, 8, 13, 16, 21, 24, 32,
39–40, 42, 53, 55, 70, 188
- I**
idiosyncratic utterances 18,
30–31
immersion xvii, 125, 163, 167,
180, 215, 239–240
implicational 39, 46, 54–55, 58,
69, 71–73, 81, 83–84, 86, 91,
96, 140, 156, 165, 184, 188
individual treatment 194–196,
198–199, 201–203
inference 223–225, 228
influencing factors 164, 185
initial state xv, 4, 12–13, 79–82,
85, 88, 94, 96–97, 240
interfaces xvii, 135, 157–158
interlanguage x, 14, 24, 39, 85,
95–96, 121, 127, 140, 142–145,
147, 155–157, 184, 187, 189, 194,
196–199, 202–203, 208, 240
interlanguage development x,
39, 121, 127, 140, 142, 144, 147,
155, 157, 197, 203, 208
- L**
L1 transfer xi, 79, 97
L2 contact duration 175,
181, 188
L2 intensity 166–167, 171,
175–179, 181, 185, 188
L2 transfer 79–80, 86, 88–89,
96–97, 240
language training ix, 105, 202
learnability xi, 122, 125–126
lexical mapping hypothesis 10,
39, 42, 45, 47
lexical verbs 15, 18–19, 21–23,
25–26, 28, 95
linguistic profiling xi, xvii,
135, 142, 145–146, 158, 163, 240
listening comprehension xvii,
207, 211, 215–216, 219
- M**
mental grammatical
system xi, xv, 3–4,
12–13, 31
mental representation 220, 227
Multiple Constraints
Hypothesis xv, 3, 12
- N**
negation 70–72, 80, 87–90, 93,
95–97, 102, 199
non-transfer approach 97
NP agreement 70–71
- O**
obligatory context 107–108,
111, 143, 188–189
- P**
parametric features 104
past tense xvi, 70–71, 101–117,
187–188
primary school xv, xvii,
14, 126, 128, 131, 163–164,
166–167, 169–170, 182–183,
186–189, 239–241
prior L2 experience 171, 181

- processability 12, 17, 38, 65–66,
69, 71, 74, 75, 76, 79, 80, 91,
97, 117, 126, 165, 189
- Processability Theory xv, 3, 30,
35, 38, 46, 58, 65–66, 76, 79,
80, 102, 110, 123, 135, 136, 140,
142, 143, 156, 163, 193, 202, 208
- proficiency rating xvii, 135
- prominence hypothesis 39–40
- R**
- Rapid Profile xvii, 135, 140,
142–148, 151, 154, 164, 175,
193–195, 198, 201
- regular verbs 101–113, 107,
116, 117
- relational hierarchy 42
- S**
- Specific Language
 Impairment xvi, 65–69,
 71–76
- SLI *see* Specific Language
 Impairment
- strongest L2 80, 89–90, 97
- Subject-verb inversion 66,
69–74
- subordinate clause 39,
69–74, 141
- suppliance 102, 107–111, 115
- Swedish xi, xvi, 48, 66, 68–71,
73–76, 79, 81, 83–88, 91–97,
141, 144, 239
- T**
- task-based language
 teaching 111, 121–126, 131, 136,
 193–194, 195, 196, 203, 240
- TBLT *see* task-based language
 teaching
- tasks with a developmentally
 moderated focus on
 form xi, xvi, 121–122,
 126–128, 130–131,
- Teachability Hypothesis
 122–123, 127
- test construct 208, 214, 221–222,
 226–227, 229, 231–232
- test-taking strategies 214, 219,
 226, 232
- thematic hierarchy 5–6, 42, 44
- thematic roles 5–7, 18–19, 22,
 25, 28, 39, 42, 44, 53
- timing of instruction 122–123
- V**
- V2 70–71, 79, 81–82, 84–89,
 91–97
- validity xvii, 80, 136, 146, 156,
 195, 207–212, 214, 220, 222,
 225–227, 230–232
- variable marking 101, 103,
 105, 116,
- variability xvi, 102, 184, 212,
 229, 232
- verbal protocol analysis 219
- vocabulary size xv, 35, 37,
 46–47, 49–52, 54–59, 62
- vocabulary size test 35,
 46–47, 49
- VP agreement 70, 72, 74–75
- vulnerable structures 68, 73–76
- W**
- Wh-questions 10, 36, 40–41,
 49–52, 56, 58
- Y**
- Yes/no questions 49

This edited volume brings together the work of a number of researchers working in the framework of Processability Theory (PT), a psycholinguistic theory of second language acquisition (SLA) (Pienemann 1998; 2005). The aim of the volume is two-fold: It engages with current issues in both theory development and theory application and focuses on theoretical developments within the framework of PT as well as issues related to second language teaching and assessment. In coordinating approaches to addressing both theoretical and applied aspects of SLA, this volume aims at bridging the gap between theory and practice. It also reflects the richness of debate within the field of PT-based research. The volume is intended for postgraduate students, SLA researchers as well as language teachers.

ISBN 978 90 272 0305 2



9 789027 203052

John Benjamins Publishing Company