ISSN 2074-9856 (print)
ISSN 2074-9864 (online)

VIENNA ENGLISH WORKING PAPERS

VOLUME 21 2012

INTERNET EDITION AVAILABLE AT:
HTTP://ANGLISTIK.UNIVIE.AC.AT/VIEWS/

CONTENTS

LETTER FROM THE EDITORS

DEAR READERS,

Because we have quite profound news to share with you at the end of this editorial, we would like to begin *in medias res*, and sketch out the research you find collected in this year's issue of VIEWS:

In the opening contribution, Anita Santner-Wolfartsberger invites us to approach a 'classic' of interaction research, namely Sacks, Schegloff and Jefferson's "A simplest systematics for the organization of turn-taking for conversation" (1974) from a new perspective. Although some readers might groan at this point that they will be faced with – yet another! – paper on turn-taking, we are confident that our author's piece indeed thrusts new light on this much-researched topic by focusing on the specificities of turn allocation in interactions involving more than two participants. Carefully unearthing the complexities of turn-taking organization in groups described by Sacks et al., Anita Santner-Wolfartsberger draws our attention to the notion of the *party*, which has hitherto been interpreted as being synonymous with a *speaker*, a fact that applies to dyadic, but not necessarily to group interactions.

Arne Lohmann in "A processing view on order in reversible and irreversible binomials" addresses the question of which factors influence the ordering of the elements in phrases such as *bread and butter* (irreversible) and *butter and milk* (reversible). Based on a rich multifactorial analysis of extensive sets of both types of data, the author proposes that "ordering in both classes can be explained via properties of the processing system, with irreversibles representing 'fossilized processing preferences'". The paper thus establishes a link between descriptive corpus-linguistic methods on the one hand, and a psycholinguistically-informed view of language structures on the other, suggesting that such a view is a promising route of research in the analysis of idiomatic constructions more generally.

The final contribution by Susanne Sweeney-Novak on "The Vienna English Language Test (VELT)", a multiple-choice test developed and used at our department to monitor access to the language competence courses, shows how teaching practice and research interests of our faculty align for mutual benefit. The article discusses the steps that were taken to ensure a scientifically sound and fair process of competence assessment for English language students. With up to 800 students of varying competence levels enrolling each semester at the department, the VELT offers a fast and easily administered way of assuring the required minimum competence level of B2 according to the Common European Framework of Reference, tailor-made to

the needs and requirements of our department, thus benefitting lecturers and students alike by providing for relatively homogenous groups of learners in our language competence courses.

While we hope that the above 'sneak previews' will make you want to skip ahead to the articles right away, we nevertheless need to return to the piece of news announced at the beginning of this editorial. One year ago, the editorial board decided on a slightly adapted publishing schedule for VIEWS, with articles appearing online as soon as they had undergone the review and revision process. However, this in turn raised the question of what the added value of our print issue was, and we ultimately decided that, in this day and age, the only honest answer was: not very much. In consequence, we have decided to focus our work and resources on the scientific review process and the publication of first-rate (online) articles in the future, and to forgo the added editorial burden and financial strain of the production of a print journal.

While this decision might be a momentous one, we in no way view it as a sad occasion.[1] Still, we would like to use this opportunity to thank all the people involved in VIEWS over the years – the founders, the contributors, the members of the editorial board and the various acting editors, the administrative staff who supported the production of the print issue, and – of course – all our subscribers and readers. At the same time, we would like to wish 'our journal' VIEWS, which has contributed so much to scholarly discussion at our department and beyond over recent years, and indeed decades, many equally successful years as an online(-only) publication!

THE EDITORS

1 To underline this point, we would like to mention that the next couple of days will see the publication of the first two online-only articles of VIEWS 22 (2013). Indeed, they might already be online by the time you are reading these lines (cf. anglistik.univie.ac.at/views/).

Parties, persons and one-at-a-time: Some fundamental concepts of conversation analysis revisited

Anita Santner-Wolfartsberger, Vienna *

1. Introduction

Spoken interaction is without doubt one of the most central social activities that human beings engage in with one another. As Zimmermann and Boden (1991: 3) put it:

> *Talk is at the heart of human existence. It is pervasive and central to human history, in every setting of human affairs, at all levels of society, in virtually every social context.*

Talk can thus justly be called "the phylogenetic and ontogenetic habitat of natural language" (Ford, Fox and Thompson 2002: 4) and it should not come as a surprise, then, that as a central social activity talk has attracted the interest of scholars from not only linguistics but various other disciplines, such as anthropology, sociolology, psychology and even the natural sciences. In light of this eminently social nature of talk it is small wonder that one of the most influential theoretical frameworks in the study of spoken interactions is an approach that originated as a branch of sociology. It was developed mainly by Gail Jefferson, Harvey Sacks and Emanuel Schegloff and came to be known as conversation analysis (henceforth CA). The present paper is concerned with some of the theoretical tenets of CA, in particular with the

* The author's e-mail address for correspondence: anita.santner-wolfartsberger@univie.ac.at

conversation analytic work carried out on the interactional practice of turn-taking.

The motivation for writing this paper stems from the difficulties I had in applying these tenets, most notably the CA premise that interactants take turns one-at-a-time, to the analysis of English as a lingua franca (ELF) data. These data consist of audio-recorded interactions among seven individuals of various first language backgrounds, who use English (or more specifically ELF)[1] as their shared medium of communication. The data hence constitute not only instances of group interactions, but also instances of intercultural communication. It is not this feature, however, that will take center stage in this paper, but the problem of accounting for the dynamics of group interaction by reference to CA turn-taking principles.

While CA has traditionally been an approach associated with monolingual (mostly American English) language use[2], Schegloff himself does not "see that there is anything, in principle, that has to be different from other work in CA" when studying 'non-native' talk (Schegloff interviewed by Wong & Olsher 2000: 113). It is therefore legitimate – and even corresponds with recent research developments in the field – to put conversation analytic principles to the test by applying them to intercultural and/or lingua franca data. This paper, however, does not include an empirical discussion of ELF turn-taking, but rather highlights what I perceive as some theoretical inconsistencies in the conversation analytic turn-taking framework that persist regardless of the nature of the data studied.

2. Basic tenets of Conversation Analysis

Conversation analysis has its roots in the works of the American sociologist Harold Garfinkel, who in the 1950s developed a new sociological discipline that he termed 'ethnomethodology'. Garfinkel wished to shift the focus of sociology away from purely quantitative analyses. Instead, he emphasized the study of 'ethnic' (i.e. the participants' own) strategies to interpret social interaction by means of commonsense knowledge and practical reasoning. In Garfinkel's own words:

---

1 For a comprehensive discussion of English as a lingua franca as a reconceptualization of English see the growing body of relevant literature, among many more e.g. Archibald, Cogo & Jenkins (2011), Mauranen & Ranta (2009), and Seidlhofer (2011).

2 However, over the last fifteen years or so, CA seems to have developed an interest in non-native or multilingual settings, as illustrated by literature on CA and Second Language Acquisition research (e.g. Markee 2000, Seedhouse 2005) or the organization of a colloquium on *New Directions in Conversation Analysis Research on L2* at the AAAL conference in March 2012.

> *Ethnomethodological studies analyze everyday activities as members' methods for making those same activities visibly-rational-and-reportable-for-all-practical-purposes, i.e. 'accountable', as organizations of commonplace everyday activities. (Garfinkel 1967: vii)*

The ethnomethodological strand of sociology hence sought to "describe the methods that people use for accounting for their own actions and those of others" (Hutchby and Wooffitt 1998: 31). Social interactions were thought to be "meaningful for interactants" and to display a "natural organization which is both discoverable and describable" (Gramkow Andersen 2001: 26).

Conversation analysis applies these ethnomethodological principles to spoken language data. Its focus is not on social interaction in general, but on talking in interaction, which is studied in the form of audio or video recordings of naturally occurring talk. Despite its name, CA is not only concerned with 'conversations' in the sense of informal small-talk, but "extends to the study of talk and other forms of conduct (including the disposition of the body in gesture, posture, facial expression, and ongoing activities in the setting) in all forms of talk in interaction" (Schegloff et al. 2002: 3). Therefore, while "ordinary conversation has a 'baseline' status in CA" (Seedhouse 2004: 1), this does not mean that non-conversational data cannot be studied using CA methodology.[3]

Conversation analysis is an interdisciplinary analytical framework situated "at a point where linguistics and sociology (and several other disciplines, anthropology and psychology among them) meet" (Schegloff 1991: 45). The main objective of CA is

> *to take singular sequences of conversation and tear them apart in such a way as to find rules, techniques, procedures, methods, maxims [...] that can be used to generate the orderly features we find in the conversations we examine. (Sacks 1984: 411)*

As Ford, Fox and Thompson (2002: 4) point out, in contrast to schools of linguistics which saw – or continue to see – naturally occurring talk as "a messy, derivative, and flawed form of language" not worthy of inquiry, proponents of CA argue that spoken interaction is a concerted activity that reflects and at the same time creates social order.

What is characteristic of conversation analysis is its orientation towards the significance of practices and rules for the interactants themselves. CA

3 However, attention needs to be payed to the fact that, according to Sacks et al. (1974: 701), non-conversational forms of talk, such as (formal) meetings, debates, interviews, or courtroom discourse, constitute separate "speech exchange systems" displaying differences in their turn-taking systems.

concentrates on local, sequential developments (Zimmermann 1988: 406) and on participants' orientation towards these developments as the interaction unfolds. Conversation analysis of talk-in-interaction is therefore

> *constrained by and focused on the participants' actual communicative activities, the finesse with which those activities are produced, and the demonstrable significance of those activities for the participants themselves. (Wooffitt 2005: 210)*

The key question conversation analysts should pose when studying a given stretch of talk can thus be summed up as *why that now?* (cf. Schegloff & Sacks 1973) or, in other words "[w]hat is getting done by virtue of that bit of conduct, done that way, in just that place?" (Schegloff et al. 2002: 5). Because this is the central issue for the interactants, it should also be the central issue for the analyst(s) (ibid.).

Communicative activities which constitute typical areas of conversation analytic inquiry include the allocation and construction of turns at talk, conversational openings and closing, other- and self-initiated repair sequences, topic management and preference organisation (Gramkow Andersen 2001: 27).[4] Among these research areas, the conversation analytic inquiry into the mechanisms that underly the allocation of speaking turns is probably CA's best known and most influential contribution to linguistics in general. In particular Sacks, Schegloff and Jefferson's (1974) account of the systematics of turn-taking for conversation has become canonical reading. Although some of its assumptions have been challenged in the past, its basic principles and terminology remain to date a standard point of reference in linguistic research on spoken discourse.

2.1 The conversation analytic turn-taking model of Sacks et al. (1974)

Turn-taking in linguistics can very generally be defined as "the process through which the party doing the talk at the moment is changed" (Goodwin 1981: 2). As such the phenomenon of turn-taking is of particular interest for those branches of linguistics studying spoken interactions. Although Erving Goffman coined the present day interpretation of turn-taking already in the 1950s[5], linguists did not pick up on this interest in turn-taking until the late 1960's, when the issue was addressed by Harvey Sacks in his *Lectures on Conversation* (Sacks 1964-72 [1995]). Some years later, Sacks et al. (1974) still noted a distinct lack of empirical research into turn-taking and observed that "no systematic account [was] available" (op. cit. 698). Their paper finally

4 See also Heritage 1985 for a more detailed list of relevant areas of CA research.

5 See Goffman (1955): "On face work".

provided just that, which probably explains the tremendous and long-lasting influence of the article.

Sacks et al.'s (1974) turn-taking framework assumes a space of interaction accessible to all participants in the conversation to the same extent. This interactional space is called the 'floor'. Participants in the interaction take turns in occupying the floor by uttering their contributions to the conversation. These contributions are called 'turns' and are

> *usually conceptualised in conversation analysis as stretches of talk by participants in verbal interaction which are concluded when the ongoing speaker is heard and/or seen to have completed what s/he was doing and then moves from the role of speaker to that of hearer. (Watts 1991: 37)*

Thus, when the current speaker finishes her turn, the now vacant position of speaker is taken up by the former listener and a new turn by a new speaker is begun. As Sacks et al. (1974: 700) argue, such turn transitions occur rapidly enough to avoid prolonged silence or gaps, but late enough to involve little or even no overlapping speech. This results in the fact that "overwhelmingly, one party talks at a time" in a single conversation (Sacks et al. 1974: 699, 706; see also Schegloff and Sacks 1973: 293).

According to Sacks et al. (1974), what makes it possible to achieve such smooth turn-transfer is the design of the turn-taking system that consists of various components, the first of which Sacks et al. (1974) have termed the "turn-constructional component" (op. cit. 702) of the system. By this Sacks et al. understand the fact that speakers may design their turns as various "unit-types", such as lexical, phrasal, clausal or sentential constructions. This allows the hearer(s) to project the unit-type under way and predict the likely completion point of the current turn before the speaker has actually reached it. The projectability of turn constructions is of particular relevance for turn-taking because possible turn completion points constitute so called transition-relevance places (TRPs), around which speaker shift takes place.

The second component of the system, which Sacks et al. call the "turn allocation component" (op. cit. 702), comprises two basic groups of turn-transfer: i) those cases where the next turn is allocated by the current speaker selecting a next speaker and ii) those cases where one of the listeners self-selects as next speaker (ibid.). The two components, the 'turn-constructional component' and the 'turn-allocation component', are complemented by "a basic set of rules governing turn construction, providing for the allocation of a next turn to one party, and coordinating transfer so as to minimize gap and overlap" (op.cit. 704). Sacks et al. formulate these rules as follows:

> *(1) For any turn, at the initial transition relevance place of an initial turn constructional unit:*

> *(a) If the turn-so-far is so constructed as to involve the use of a 'current speaker selects next' technique, then the party so selected has the right and is obliged to take next turn to speak; no others have such rights or obligations, and transfer occurs at that place.*
>
> *(b) If the turn-so-far is so constructed as not to involve the use of a 'current speaker selects next' technique, then self-selection for next speakership may, but need not, be instituted; first starter acquires rights to a turn, and transfer occurs at that place.*
>
> *(c) If the turn-so-far is so constructed as not to involve the use of a 'current speaker selects next' technique, then current speaker may, but need not continue, unless another self-selects.*
>
> *(2) If, at the initial transition-relevance place on an initial turn-constructional unit, neither 1a nor 1b has operated, and, following the provision of 1c, current speaker has continued, then the rule-set a–c re-applies at the next tansition-relevance place, and recursively at each next transition-relevance place, until transfer is effected. (Sacks et al. 1974: 704)*

According to Sacks et al. (1974), the above rule-set, the turn-constructional component and the turn allocation component work together in "organizing transfer exclusivley around transition-relevance places" and so "provide for the possibility of transitions with no gap and no overlap" (op. cit. 708).

The great achievement of Sacks et al.'s systematics for turn-taking lies in the fact that it uncovered the precision and orderliness underlying such a mundane activity as ordinary conversation. In this regard, as Oreström (1983: 29) points out, the framework represents "a valuable achievement as it contributes to the understanding of how conversation is systematically ordered". However, as he continues, although Sacks et al.'s work "is a great step towards such an understanding", the model has its "obvious limitations" (ibid.). One focal point of the critique has very often been CA's conceptualization of simultaneous speech as a violation of the one-at-a-time principle and thus a kind of 'noticeable' linguistic behaviour that requires repair.

2.2 Dealing with simultaneous speech in CA

According to Sacks et al., speaker shift may only occur at or around transition-relevance places in order to be in conformity with the turn-taking system (op. cit. 706, 708). As any interactant has the right (and obligation) to continue the current turn until its first possible completion point, turn entry by another speaker before this first transition-relevance place is prone to result in both speakers talking at the same time. This violates the basic principle of 'one-at-a-time' and will therefore, according to Sacks et al., be perceived by the interactants as a violation of turn-taking principles. In cases of such

'erroneous' turn-taking, so called "repair mechanisms" (op. cit. 701) come into play, which in this case would consist of one of the speakers stopping their turn prematurely:

> *Thus, the basic device for repairing 'more than one at a time' involves a procedure which is itself otherwise violative in turn-taking terms, namely stopping a turn before its possible completion point. (op.cit. 724)*

This interpretation of overlap as a violative aberration within the turn-taking system that requires repair reflects the popular belief that the more often a speaker 'interrupts' others, the more s/he can be seen as dominating the interaction. The conceptualization of any overlap as 'interruption' in early CA studies invited a series of investigations on gender differences (e.g. Eakins and Eakins 1976, Leet-Pellegrini 1980, Murray and Covelli 1988) in the frequency of the (non)use of simultaneous speech in the 1970s and 1980s, which will not be discussed further here.[6] It suffices to state that, as Watts (1991) points out, for quite some time, simultaneous speech was simply equated with uncooperative interruptions before scholars acknowledged the existence of unproblematic overlaps and recognized that "simultaneous speech can in certain cases be a sign of cooperation and is not necessarily understood as an interruption by the interactants" (Oreström 1983: 147).

Today, the existence of unproblematic overlap is a widely acknowledged feature, both within and beyond the CA research community. Ample research (e.g. Ford & Thompson 1996: 157-164; Gramkow Andersen 2001; Lerner 1993, 1996, 2002, 2004; Meierkord 2000; Tannen 1984 [2005]) has demonstrated that, far from being a blatant or random violation of turn-taking rules, "speakers' production of simultaneous talk is ordered, consequential and functional, precisely in relation to the projectable linguistic and social unfolding of turns" (Ford, Fox and Thompson 2002: 8). The discrepancy between the theoretical assumption of one-at-a-time in their model, and the reality of simultaneous talk in interactions was also acknowledged by Sacks, Schegloff and Jefferson themselves, as can be seen by the continued work of Jefferson and Schegloff on overlap for decades after the appearance of their article in 1974 (e.g. Jefferson 1983a, 1983b, 1984, 1986, 2004; Schegloff 1996, 2000, 2002).[7] But while the frequent ocurrence of overlap has led some

---

[6] For a more thorough discussion of this issue refer to Wolfartsberger (2011a).

[7] Schegloff (2000), for instance, has continued to work on the resolution of overlap by interactants according to the framework proposed by Sacks et al. but has "for more than 25 years withheld and extended a formal writeup of this work in the hope of expanding the data base in which it was grounded, especially with respect to video data, and in the hope of the substantial and continually growing literature of this much focused-on topic" (op. cit. 46).

scholars outside the CA community to question and criticize conversation analytic concepts like turn, floor, or the principle of one-at-a-time (e.g. Edelsky 1981, Meierkord 2000), Schegloff (2000, 2002) defends the general validity of these concepts and argues that there are certain exceptions to the one-at-a-time principle.

In an article on overlap resolution, Schegloff (2000) acknowledges the existence of unproblematic overlap and addresses certain aspects of the turn-taking organization with regard to overlap which have remained unclear due to "several underspecifications in our previous account of that organization" (op.cit. 42). Schegloff investigates how interlocutors deal with simultaneous speech once it occurs, his goal being to develop "a model of an 'overlap-resolution device'" (op. cit. 4). In doing so he excludes four types of overlapping talk from his materials which are not taken as problematic by the interlocutors because "the simultaneous speakers do not appear to be contesting or even alternative claimants for a turn space" (ibid.). These four exceptions would be i) backchannel utterances, which Schegloff terms 'continuers', ii) terminal overlaps, iii) choral forms of talk like laughter, collective greetings or congratulations and finally iv) utterances that constitute conditional entry into a turn (cf. Lerner 1996) only, such as help in word search situations. However, even if one excludes these cases of "unproblematic overlap", cases persist of simultaneous speech which the turn-taking framework proposed by Sacks et al. (1974) cannot satisfactorily explain.

It is not surprising, then, that the one-at-a-time principle and the line taken by Sacks et al. on simultaneous speech have not gone uncontested in the literature. Criticism mainly focuses on the fact that Sacks et al. (1974) claim universal validity for the contentions outlined in their article. This becomes obvious when they argue that the rules for the organization of turn-taking they describe hold true "in any conversation" (Sacks et al. 1974: 700). It is exactly this claim of universal validity that is one of the aspects most often criticized about Sacks et al.'s turn-taking framework, which is seen by many scholars as unable to cope with (cross)-cultural variation (e.g. Agrawal 1976, Kilpatrick 1990, Larson & Dodds 1985, Moerman 1987, Reisman 1974; see also Meierkord 2000 and Gramkow Andersen 2001 for English as a lingua franca). The incapability of Sacks et al.'s (1974) model to account for variation in turn-taking, especially with regard to overlaps and interruptions, has led scholars like Makri-Tsilipakou (1994: 403) to even diagnose a '"one at a time' bias of the white American turn-taking model" which, in her opinion, "has [...] been shown to be inoperative across different cultures".

While this is no doubt a relevant issue, culturally or ethnically motivated variation is not the focus of this paper. Rather, what is of interest is a second strand of critique coming from scholars dealing with group interactions (e.g. Edelsky 1981, Kerbrat-Orecchioni 2004) who found it hard to apply Sacks et al.'s framework to interactional data involving more than two participants although its authors claim that "the system is compatible with different numbers of participants from conversation to conversation" (Sacks et. al.: 712). The question of whether the number of interlocutors impacts in any way on the organization of turns at talk is therefore a relevant issue in need of further exploration.

3. Turn-taking in Multiparticipant Interaction

3.1 Hidden assumptions: a 'dyadic dictat'?

If one searches the literature for accounts of turn-taking in group interactions one soon notices that "research has addressed fairly little the question of the specific ways in which multiparty conversation differs from dyadic conversation" (Kangasharju 1996: 296). As Goffman (1981) points out:

> *Traditional analysis of saying and what gets said seems tacitly committed to the following paradigm: two and only two individuals are engaged together in it, [...] the two-person arrangement being the one that informs the underlying imagery we have about face-to-face interaction. (op. cit. 129)*

Similarly, Kerbrat-Orecchioni (2004: 2) also diagnoses a "deep-rooted tendency to associate interaction with interaction between two people" and to consider dyadic interactions "as the prototype of all forms of interaction". This tendency is even more surprising given the fact that, according to Kerbrat-Orecchioni, in any society dyadic exchanges tend to be a minority of all face-to-face interactions (ibid). Nevertheless, as Kerbrat-Orecchioni argues, the image of dyadic interactions as the prototypical form of verbal communication seems to be a popular belief:

> *Dyadic communication is widely thought to be the communicative situation par excellence – not only by linguists, semioticians, psychologists, and communication theoreticians, but also by 'the man on the street'. (op. cit. 1)*

For some time now this "dyadic diktat" (op. cit. 2) has been criticized by scholars specializing in the analysis of spoken discourse. Levinson (1988: 222-223), for instance, argues that this "bias towards the study of dyadic interaction" in the study of verbal exchanges acts as a "straightjacket" which limits the research foci of studies and confines analysts to certain situations

and cultures. Hence, he regards the "dyadic triumph" (ibid.) as ethnocentric and in no way beneficial to the study of verbal interactions.

The scope of situations in which interactions between more than two participants can be witnessed and examined is thus a vast research field which has been tackled by only a few scholars so far. The neglect of turn-taking in group interactions in conversational research is quite astounding, particularly as a major proponent of CA has drawn attention to the issue from the beginning. As early as 1967, Sacks in his *Lectures on Conversation* pointed out that "attention has to be payed directly, independently, to multi-party conversations, and there's ways in which they could be much more interesting" than dyadic conversations, which are "much blander" (op. cit. 533). Sacks (1967) therefore urged his audience to view group interactions as

> *a distinct phenomenon; and order of facts; something to be investigated in its own terms, and not merely – as I think a good deal of my discussion of that phenomenon treated it – as a variant* ***off*** *[sic!] two-party conversation" (op. cit. 523, original emphasis).*

However, Sacks' call for more research went unheard. Almost 30 years later, Schegloff (1995) still felt the need to draw attention to the issue. For him, the preferential treatment of dyadic talk-in-interaction is to be seen in connection with the historical development of academic disciplinces such as linguistics and social psychology:

> *With the so-called 'linguistic turn' in studies of the domain which was previously the prerogative of social psychology, [...] occasions which were dialogic, i.e., composed of two participants came often to be referred to generically as 'conversation' or 'interaction.' [!] (Schegloff 1995: 31)*

In summary it can be said that, while there is theorizing about turn-taking in multi-person interactions in CA to a certain extent, most empirical work centers on dyadic speaker situations. Theoretical claims on multi-person talk, however, need to be backed up by empirical research, which holds the potential to enhance our knowledge on the underlying systematics governing the organization of face-to-face communication in groups.

3.2 Turn-taking in group interactions

A notable exception to the general neglect of turn-taking in multi-person settings is the *Groupe de Recherches sur les Interactions Communicatives*, a team of French researchers based in Lyon, who have carried out extensive research on interactions involving three or more participants. The research team has coined the term *polylogue* (or *polylogal*) for "all communicative

situations which gather together several participants, that is, real live individuals" (Kerbrat-Orecchioni 2004: 3).[8] Polylogues can include "a theoretically infinite number of participants" (op.cit. 4), as is the case, for example, in internet newsgroups (e.g. Marcoccia 2004). However, as Kerbrat-Orecchioni points out, "[b]eyond four participants, the problems of describing the interaction increase dramatically" (op. cit. 7).

One of the most obvious reasons for this is the possibility of "schism of one conversation into more than one conversation" (Sacks et al. 1974: 713), i.e. the development of parallel conversations, in situations involving four of more interlocutors. Turn-taking in multi-participant interactions therefore involves the issue of regulating the "distribution of opportunities to talk among the several participants – including at times the forced draft of ones who appear in danger of drifting into schisms" (Schegloff 1996: 21).

However, even if schisming does not occur, multi-participant interactions still pose a challenge to analysts. One of the prime reasons for this is their "variability in alternation patterns" with regard to turn-taking, which manifests itself in a "lack of balance in floor-holding, violations of speaker-selection rules, and interruptions and simultaneous talk" (Kerbrat-Orecchioni 2004: 4). This heightened variability in the realm of turn-taking can simply be explained by the increased number of interlocutors, which amplifies the complexity of turn-taking mechanisms and sequential rules governing speaker shift. For instance, as Sacks et al. (1974: 712) have pointed out, "for two parties, the relevant variability is not differential distribution of turns (given that they will have alternating turns), but differential turn size". This means that

> *[i]n two-party conversation, a current non-speaker can pass any given transition-relevance place which is non-obligatory (i.e., where current selects next technique has not been used) with full assurance of being 'next speaker' at some point; but with three or more parties, this is not assured. (ibid.)*

Put differently, while distribution of speakership in dyadic interactions always follows a pattern of A-B-A-B ..., an interaction with three participants is not necessarily characterized by an A-B-C-A-B-C ... pattern (nor are interactions with more than three interlocutors). What in fact happens, is that

> *one party will be responding to the immediately prior turn and the immediately prior turn will in its turn be responding to its immediately prior turn. And the*

8 The reason why they stress that a polylogue has to involve several "real live individuals" lies in the terminological confusion regarding the term 'multi-party conversation' which is used in most of the conversation analytic literature and which will be the focus of section four.

> *majority of turns produced in this way will be designed with the producer of the turn to which it is a response as its intended recipient. This will inevitably create difficulties for other participants [...] who want to break into the conversation. (Langford 1994: 108)*

Through its sequential rules for allocating speakership (see section 2.1), Sacks et al.'s framework hence favors the 'just prior to current' speaker to be selected as next speaker (because the 'current speaker selects next' option has priority over other selection techniques). This turn order bias, as Sacks et al. point out, is only operative in group interactions:

> *In two-party conversation, the two speakers to whom the rule-set refers, and for whom the turn-order bias works, comprise all the parties to the conversation, and it is not in point to speak of a turn-order 'bias'. The 'last as next' bias, however, remains invariant over increases in the number of parties — and, with each additional increment in number of parties, tends progressively to concentrate the distribution of turns among a sub-set of the potential next speakers. With three parties, one might be 'left out' were the bias to operate stringently; with four parties, two would be 'left out', etc. (Sacks et al. 1974: 712)*

Therefore, as Sacks et al. (1974: 708) point out, "while turn order varies, it does not vary randomly" in group interactions. Consequently, any participant currently not in the role of speaker who is interested in occupying the floor next will pay very close attention as to when the current speaker is coming to a possible completion, and hence a TRP, in order to be able to self-select before the current speaker selects a next speaker. At the same time, a current speaker who wants to choose the next speaker will have to select this next speaker before a possible transition relevance place in order to prevent another participant from self-selecting and taking over the turn at that point (op. cit. 713). The sequential rule of "first starter goes" will thus result in "pressure for minimization of turn size, distinctively operative with three or more parties" (Sacks et al.: 713). This pressure, resulting from the increased number of potential next speakers in multi-participant interactions, heightens the probability for current non-speakers to interfere with the current speaker's turn. Not surprisingly, Kerbrat-Orecchioni (2004) finds that

> *[t]he frequency of interruptions and simultaneous talk as well as the variety of ways in which these are carried out, increases in trilogues, and a fortiori in multi-participant interactions. (op. cit. 5)*

However, this does not mean that polylogues are per se characterized by more competitive turn-taking behavior than dyadic interactions. Kerbrat-Orecchioni refers to Müller (1995), who studied discussions among eight French students and reports an extensive use of overlap, mostly three or even four participants

speaking at the same time.[9] Despite this first impression of "unbearable cacophony", a detailed analysis "reveals the concerted organizations of these interruptions, which more often than not have a collaborative function" (Kerbrat-Orecchioni 2004: 5). The use of the term 'interruption' for collaborative overlaps in this quotation is thus slightly misleading and testifies to the terminological confusion regarding the term 'interruption' that prevails in much of the literature on overlap. However, it would go too far to embark on a terminological debate about this matter here.[10]

For the purpose of this paper it suffices to say that the number of participants in a given interaction impacts, in various ways, on the allocation of speaking turns. While a greater number of interlocutors does not necessarily result in a more competitive way of talking, the likelihood for overlaps to occur is indeed increased, and so is the complexity of mechanisms governing turn design and speaker shift. However, this complexity does not only pose a challenge to the analyst, but also to the interlocutors:

> *[F]or the analyst, the functioning of trilogues is in all regards more complicated to describe than that of dilogues [sic!]. For the participants themselves, the more numerous they are, the more delicate conversational activities become. Speakers must take all their recipients into account to some degree, and the recipients themselves are intrinsically heterogeneous due to differences in status, knowledge, expectations, objectives, etc. (Kerbrat-Orecchioni 2004: 6)*

This also means that the conversation analytic concept of 'recipient design' is characterized by increased complexity, as speakers in certain situations may have to aim at multiple recipient design (ibid). In summary it can be said that the "leitmotiv", as Kerbrat-Orecchioni calls it, is "that of the extreme complexity and flexibility of polylogal organizations" which "would be enough to discourage any researcher" (op. cit. 20). In fact,

> *polylogues have an organization which is so mobile and so changeable that observing them at a t1 point in time can never provide a representative picture of the whole. (ibid.)*

---

9 Müller's observation runs counter to Schegloff's (1995) claim that overlap in interaction usually means two (and not more) speakers overlapping with one another. Consider:

> "Not only is it empirically the case that more than one speaker at a time is almost always two speakers at a time; it is also the case which requires no more than two – the case where two speakers are speaking to each other – which is the general case of overlap, the one with which inquiry must begin. Whereas for turn-taking in general 'two' is precisely not the general case, for overlap it precisely is." (Schegloff 1995: 40, original emphasis)

Clearly, more empirical work on this issue is necessary.

10 But see Wolfartsberger (2011a, in. prep.) for a discussion of the issue.

It is therefore not surprising that researchers who tried to analyze non-dyadic interactions within the traditional CA framework have encountered problems (e.g. Edelsky 1981; Meierkord 2000; Kerbrat-Orecchioni 2004; Watts 1991, Wolfartsberger 2011b) and proposed amendments to Sacks et al.'s model. The reason for these difficulties are partly also to be found in the framework itself.

In principle Sacks et al.'s (1974) turn-taking model is designed to accommodate any number of interlocutors, not just two, and in any kind of conversational setting.[11] This is exemplified by the following observation which, according to Sacks et al. (1974: 701), applies to any conversation:

> *(10) Number of parties can vary (cf. §4.10).*

The assumption then is that, just as the turn-taking model can account for conversations of various lengths, it can also account for conversations among a varying number of participants, including exit and entry of participants during a single conversation (op. cit. 712). However, as we have already seen, a closer reading of Sacks et al.'s article reveals that there are certain limits as to what the system provides with regard to the number of interlocutors: the sequential rules for speaker shift (outlined in section 2 above) organize only two speakers, namely 'current' and 'next'. The result of this is the 'last as next' bias discussed above, which leads to the fact that,

> *[t]hough the turn-taking system does not restrict the number of parties to a conversation it organizes, still the system favors, by virtue of its design, smaller numbers of participants. (Sacks et al. 1974: 712)*

The question hence arises what should be understood by a "smaller number" of participants: three instead of five? Five instead of ten? Equally left unclear is whether there is a maximum number of participants the system can deal with.[12] After all, instances of four or five individuals engaged in conversation are by no means rare and any viable turn-taking model must be compatible with them. But there is another issue that is even more problematic about the framework's application to group interaction, and that is the notion of 'party'.

---

11 In the original framework, Sacks et al. (1974) argue that this does not include talk in institutional settings, such as (formal) meetings, debates, interviews, or courtroom discourse, which are separate "speech exchange systems" (Sacks et al. 1974: 701) displaying differences in their turn-taking systems. But see Schegloff et al.'s (2002) take on this discussed in section two.

12 Naturally, there are limits of a purely practical and acoustic nature of how many individuals can engage in talk-in-interaction at the same time.

4. One-at-a-time: parties or persons?[13]

Much of the criticism of Sacks et al.'s turn-taking model revolves around the issue of one-at-a-time. So let us examine again what Schegloff himself has to say about the matter. At the very beginning of his article on overlap resolution, Schegloff (2000) writes that the turn-taking organization

> *is an organization of practices designed to allow routine achievement of what appears to be overwhelmingly the most common default 'numerical' value of speakership in talk-in-interaction:* ***one party*** *talking at a time. (Schegloff 2000: 1, emphasis added)*

In this quote it is obvious that Schegloff (2000) understands the one-at-a-time constraint to be operating between parties. Yet, on the next page, we find Schegloff arguing that

> *[t]alk by more than* ***one person*** *at a time in the same conversation is one of the two major departures that occur from what appears to be a basic design feature of conversation, and of talk-in-interaction more generally, namely 'one at a time' (the other departure is silence, i.e. fewer than one at a time). (op.cit. 2, emphasis added)*

Clearly, here Schegloff argues that one-at-a-time is to be understood as one-person-at-a-time. If we assume that *party* is synonymous with *person*, there is nothing to say against these quotations, apart from the fact that quite a number of scholars have found evidence in their data that interlocturos do *not* always orient to this default numerical value of speakership. However, the reader of Schegloff's article might be a bit puzzled to learn in note eight of the very same article that this obviously is not the case. A party is not necessarily the same as a person, at least in group interactions, as Schegloff talks about the fact that

> *co-members of a party [...] may come to talk simultaneously because their party has been selected to speak next, but in a fashion that does not specify which of them is to do the speaking. (op.cit. 48, note 8)*

According to Schegloff, such overlaps are "not infrequent" and "systematic products of a turn-taking organization which allocates turns to parties, but not necessarily among party-coincumbents" (ibid.). Obviously, then, for Schegloff a *party* is not necessarily the same as a *person*, as a party may be composed out of several individuals acting as one party to talk-in-interaction. Though Schegloff (2000) admits that the issue has not yet been studied systematically, he refers to an earlier empirical investigation of the matter in

---

13 In this article, the term 'party' is used as Schegloff (1995) understands it and *not* symonymously with the terms 'person', 'speaker', 'participant', 'interactant', etc., all of which refer to individuals.

which overlap happens exclusively among co-incumbents of a single party (Schegloff 1995). In that article Schegloff highlights that

> *the turn-taking system as described in SSJ [Sacks, Schegloff and Jefferson 1974] organizes the distribution of talk not in the first instance among persons, but among parties. [...] on some occasions, or for some particular phase or topic sequence within some occasion of talk-in-interaction, the aggregate of persons who are, as Erving Goffman called them "ratified participants", are organized into parties, such that there are fewer parties than there are persons. (Schegloff 1995: 32-33)*

In exchanges between two interactants the number of persons usually corresponds to the number of parties. With three or more participants, this is not necessarily the case, as apparently several persons might form one party for a given stretch of talk. The conceptual distinction between *party* and *person* hence becomes only relevant if one deals with talk-in-interaction among more than two individuals. Schegloff (1995) elaborates on the issue a bit more and highlights that the mechanisms for selecting a next speaker outlined in Sacks et al. (1974) operate on the level of parties, not on the level of persons:

> *If there are multi-person parties in the interaction, the turn-taking organization does not necessarily provide for the selection of a person to speak for the party, nor does it provide procedures for doing so (aside from a procedure, or device, for resolving overlapping talk if/when it arises [...]). (Schegloff 1995: 33)*

Hence, while the turn-taking system outlined in Sacks et al. (1974) provides for the selection of a next *party*, it does not necessarily provide for the selection of an individual *speaker* for that party. All it does is to offer repair mechanisms to resolve overlap when it occurs.

For Schegloff (1995), the conceptual distinction between parties and persons is crucial in that it provides an opportunity to further defend Sacks et al.'s (1974) turn-taking model and to refute criticism brought forward against it:

> *In **assessing the adequacy of the SSJ model** of turn-taking, such considerations will be important, for without them we will not properly appreciate the character of different kinds of simultaneous talk for the participants, and therefore their different bearing on assessment of the model. (Schegloff 1995: 35, original emphasis)*

I find Schegloff's argument slightly perplexing, for various reasons. First, it is unclear to me whether 'one-at-a-time' is to be understood now as 'one-*party*-at a time' or 'one-*person*-at a time' since both phrases are used by Schegloff (2000). Second, in my opinion it is not very convincing to claim, on the one hand, that turn-taking operates between parties (not speakers), but simultaneously argue, as Schegloff does, that the underlying principle

speakers orient to when taking turns at talk is 'one *speaker* at a time', and not one *party*. The resolution of this issue, however, seems to be crucial in order to be able to investigate turn-taking and simultaneous speech in group interactions.

Schegloff also seems to twist the argument for his own benefit with regard to unproblematic overlap: In his 1995 article, he argues that "the turn-taking system as described in SSJ organizes the distribution of talk not in the first instance among persons, but among *parties*" (Schegloff 1995: 32-33, emphasis added) and uses this to explain the occurrence of overlaps in group interactions. On the other hand, five years later, Schegloff (2000) states that "SSJ's claim was that turn-taking is organized by reference to one-*speaker*-at-a-time [...]" (op. cit. 47, emphasis added), which construes overlap as violative behavior. The contradiction is, in my view, conspicuous. Also, I cannot quite see why Schegloff (1995) establishes the distinction between party and person in the first place, and then in his account of overlap resolution (Schegloff 2000) goes back to equating parties and persons, as we have seen. Finally, I cannot help wondering why, if the distinction between party and person had been so obvious from the beginning, Sacks et al. (1974) did not discuss the significance of this point and its relevance for multi-person interactions more prominently in their original framework.

Given the fact that Sacks et al.'s original framework does not mention the distinction between *party* and *person* at all and that Schegloff seems to use the terms *party* and *person* simultaneously in his later work on overlap (2000, 2002), it is not surprising that the issue seems to have escaped the attention of a large number of researchers, I dare say the majority. Goodwin, for instance, clearly equates a party with an individual person:

> *A party whose turn is in progress at a particular point in time will be called a speaker. (Goodwin 1981: 3).*

Moreover, researchers criticizing Sacks et al.'s premise of one-at-a-time have also interpreted it as pertaining to individual speakers, not parties, even after the publication of Schegloff (1995). One quote by Gramkow Andersen may serve as an exemplary illustration of this:

> *In the future we shall have to emphasize the 'one **speaker** at a time' principle somewhat less than we have done so far, in the sense that the speakers, during simultaneous speech passages, rarely orient to the fact that they are 'violating' the turn taking system. (Gramkow Andersen 2001: 157, emphasis added)*

Similarly, the terms 'multi-party' and 'multiparticipant' talk seem to be used interchangeably in the literature to mean the same thing, even though, according to Schegloff, they do not.[14] Most scholars therefore seem to use the term 'multi-party interaction' without acknowledging that in doing so they subscribe to Schegloff's understanding of 'party'.

The only instance of the distinction between person and party being taken up in the literature is found in Kerbrat-Orecchioni's (2004) account of research on polylogues carried out by herself and her colleagues. She rejects the idea that turn-taking operates between parties in Schegloff's (1995) sense. For her it is clear that "turn-taking operates per se between speakers" and that "the succession of turns is first and foremost a phenomenon which takes place betweeen individuals" (op. cit. 3), even if she admits that alignment between speakers depending on their objectives, statuses, and roles is doubtlessly important for the analysis of interaction. When talking about alignment, however, she argues that we need to distinguish between three levels: the level of speakers, the level of interactional roles, and the level of discursive roles (Bruxelles and Kerbrat-Orecchioni 2004: 110-111). Kerbrat-Orecchioni and her colleagues argue that the notion of 'party' belongs to a higher analytical level which cannot be determined on the purely structural level of speaking turns. After all, "[w]hereas the number of participants in a polylogue is an objective fact, 'parties' can only be determined from a specific point of view" (ibid.). However, even Kerbrat-Orecchioni at times seems to confuse party with individual participants, as in the following quote:

> *A coalition can also be based on active collaboration on the part of a potential ally, who comes to the aid of a particular **party** and helps to fulfill **his or her** illocutionary and argumentative goals. (op.cit. 80)*

I therefore think it is fair to say that the concept of party as potentially encompassing several interlocutors has not really caught on in the scientific literature concerned with turn-taking in spoken interaction. Nevertheless, its implications, particularly for group interactions, are considerable.

5. So why that now? Implications for further research

Though seemingly only a minor issue, the distinction between *party* and *speaker* as postulated by Schegloff (1995) is not without consequences. It can

[14] Jenks (2009, 2011), for instance, seems to be unaware of the distinction between party and participant, using the expression "multi-party interaction" in one publication and "multiparticipant" in another to describe situations involving more than two interlocutors.

hardly be denied that such a distinction would cast some, if not all, of the principles laid out in Sacks et al. (1974) in a new light. For instance, the mechanisms and rule-sets postulated there for regulating speaker change (see section two above) like the 'current speaker selects next' technique (hitherto interpreted in the literature as 'current speaker selects next *speaker*'), would have to be re-examined in order to include the possibility of selecting a next *party* consisting of several speakers. This holds true for basically all studies that have investigated turn-taking in group interactions so far and have treated parties as synonymous with individual participants. Some might say that one should think twice before doing this only because an individual researcher, even if this individual researcher is Emanuel Schegloff, differentiates between parties and participants in one article. If we pretend not to have read Schegloff (1995) and go back to equating parties with persons, so one might argue, then all is well.

Such a line of argumentation, however, ignores the fact that over the years considerable evidence has come to the fore that the conversation analytic turn-taking model with its underlying assumption of one-at-a-time (interpreted as one *speaker* at a time) cannot always explain what analysts find in their data, especially in group settings or in other-than-monolingual-English contexts. In fact, if we look at the criticism brought forward against Sacks et al. (1974), we can identify mainly two groups of researchers critical of Sacks et al.'s turn-taking framework: i) researchers dealing with group interactions and ii) scholars working with non-English or intercultural data who question the cross-cultural validity of the model. Although both criticisms revolve mostly around the principle of one-at-a-time, they have hitherto been seen as unrelated.

In my view, however, it is possible that they both stem from the same issue, i.e. the inconsistent conceptualization of one-at-a-time in relation to parties and speakers. For if we assume that the underlying principle interactants orient to is indeed one *party* at a time, not one speaker, it follows that what is a party will very likely be construed (and perceived) differently by different interactants in different situations. The notion of party might hence offer an explanation for the frequent ocurrence of overlap and the different reactions it triggers by the interactants, who sometimes orient to it as interruptive, and sometimes as collaborative, and sometimes even in both ways (cf. Tannen 1994). A distinction between *party* and *person* implies that there are two different types of overlap possible in interaction: overlap of party co-members, and overlap of speakers belonging to separate parties. As we have seen above, Schegloff (2000) understands the one-at-a-time constraint to be operating between parties, but not between speakers.

Consequently, overlap among party co-members would be in conformity with the rules set out in Sacks et al. (1974), whereas overlap across parties would constitute a violation of these turn-taking rules.

At least, that is the direction in which Schegloff (1995) seems to argue, as he contends that there is a difference between overlaps among co-incumbents of a party, and overlaps between parties:

> *Consequently, in understanding the interactional significance of simultaneous talk-in-interaction, and in appreciating its relevance of the assessment of models of turn-taking, one important discrimination will be between simultaneous talk between coincumbents of a single party on the one hand, and between separate parties on the other. (Schegloff 1995: 33)*

He also claims that in sequences where there is frequent overlap "much, and often all, of the simultaneous talk is between participants who, at the moment in the conversation, are co-incumbents of a party" (ibid.). It is not clear, however, to what extent this claim is based ony systematic empirical investigation. In any case such an argumentation suggests that overlap among co-party members is not treated as violative, whereas overlap across parties is. One wonders if the difference between an interruption and a supportive completion then should be conceptualized in the way that the intervening speaker is a member of a different party in the case of interruptions and a co-incumbent of the same party in the case of supportive completions.

The crucial question that poses itself, however, is how a party is defined. Unfortunately, there is only little information provided by Schegloff (1995: 34-35) on this crucial question. He discusses just one extract of data where four interactants are divided up into a party of "the informed" vs. a party of "the uniformed". All we learn about the concept of party in the brief discussion of this one example is that for Schegloff a *party* is not a stable concept that remains the same throughout the interaction. As "people can come and go in the course of talk-in-interaction" (Schegloff 1995: 35), it follows that even if the overall number of participants in a single conversation remains the same,

> *the number of parties into which those participants may be seen to be organized (because they see **themselves** so to be organized, and embody that stance in their conduct) can change continuously as the contingencies of the talk change, [...] (Schegloff 1995: 35, original emphasis)*

Schegloff obviously views a party as something that is constantly in flux throughout the interaction, as well as something that needs to be displayed by the interactants and – presumably – also negotiated.While the concept of party seemingly explains why not all cases of overlap are treated as violative behaviour by the participants, it raises a host of other questions that are

desperately in need of clarification. To name just a few: How is a party defined/identified by the analyst? How is it construed/identified by the interactants? Is it construed/identified by all interactants in the same way? In all cultural contexts? How does it relate to 'turn' and 'floor'? Is the floor occupied by one speaker or one party? And so forth. The introduction of *party* seems to raise more questions than it answers at the moment. Clearly, a systematic empirical exploration of the phenomenon is needed before we can try to answer any of these questions.

References

Agrawal, Arpita. 1976. "Who will speak next". *Papers in Linguistic Analysis. Department of Linguistics, University of Delhi 1*: 58-71.

Archibald, Alasdair; Cogo, Alessia; Jenkins, Jennifer. 2011. *Latest Trends in ELF research.* Newcastle upon Tyne: Cambridge Scholars Publishing.

Eakins, Barbara and Gene Eakins. 1976. "Verbal turn-taking and exchanges in faculty dialogue". In: Betty Lou Dubois and Isabel Crouch (eds.): *The Sociology of the Language of American Women*. Papers in Southwest English IV, Trinity University, San Antonio, 53-62.

Edelsky, Carole. 1981. "Who's got the floor?" In: *Language in Society* 10, 383-421.

Ford, Cecilia E., Barbara A. Fox and Sandra A. Thompson (2002): "Introduction". In: Cecilia E. Ford, Barbara A. Fox and Sandra A. Thompson (eds.): *The Language of Turn and Sequence.* Oxford: Oxford University Press.

Ford, Cecilia E., Barbara A. Fox, and Sandra A. Thompson (eds.) (1996): *The Language of Turn and Sequence.* Oxford: Oxford University Press.

Garfinkel, Harold. 1967. *Studies in Ethnomethodology*. Englewood Cliffs, N.J.: Prentice Hall.

Goffman, Erving. 1955. "On Face Work: An Analysis of Ritual Elements of Social Interaction" *Psychiatry: Journal for the Study of Interpersonal Processes* 18(3), 213-231.

Goffman, Erving. 1981. *Forms of Talk.* Philadelphia: University of Pennsylvania Press.

Goodwin, Charles 1981. Conversational Organization. Interaction between Speakers and Hearers. New York: Academic Press.

Gramkow Andersen, Karsten, 2001. *The Joint Production of Conversation. Turn-sharing and collaborative overlap in encounters between non-native speakers of English*. Aalborg: Centre for Languages and Intercultural Studies Aalborg University.

Heritage, John. 1985. "Recent Developments in Conversation Analysis." Sociolinguistics 15, 1-18.

Hutchby, Ian; Wooffitt, Robin. 1998. *Conversation Analysis: Principles, Practices and Applications*. Cambridge: Polity Press.

Jefferson, Gail. 1983a. "On a Failed Hypothesis: 'Conjunctionals' as Overlap-Vulnerable". *Tilburg Papers in Language and Literature*, No. 28, 1-33. Tilburg: Tilburg University.

Jefferson, Gail. 1983b. "Another Failed Hypothesis: Pitch/Loudness as Relevant to Overlap Resolution". *Tilburg Papers in Language and Literature,* No. 38, 1-24. Tilburg: Tilburg University.

Jefferson, Gail. 1984. "Notes on some orderlinesses of overlap onset". In V. D'Urso and P. Leonardi (eds.) *Discoure analysis and natural rhetoric.* Padua, Italy: Cleup Editore, 11-38.

Jefferson, Gail. 1986. "Notes on 'latency' in overlap onset". *Human Studies, 9*(2/3), 153-183.

Jefferson, Gail. 2004. "A sketch of some orderly aspects of overlap in natural conversation". In G. H. Lerner (ed.) *Conversation Analysis: Studies from the first generation.* Philadelphia: John Benjamins, 43-59.

Jenks, Christopher J. (2009) When is it appropriate to talk? Managing overlapping talk in multi-participant voice-based chat rooms. *Computer Assisted Language Learning*, 22 (1), 19-30.

Jenks, Christopher J. 2011. *Transcribing Talk and Interaction: Issues in the Representation of Communication Data*. John Benjamins.

Kangasharju, Helena. 1996. "Aligning as a team in multiparty conversation". *Journal of Pragmatics 26* (1996) 291-319.

Kerbrat-Orrecchioni, Catherine. 2004. "Introducing Polylogue". *Journal of Pragmatics* 36, 1-24.

Kilpatrick, Paul W. 1990: "Comprehension of simultaneous speech in conversation". Paper presented to the 9th World congress of Applied Linguistics, Thessaloniki, Greece, April 15-21, 1990.

Langford, David. 1994. *Analysing Talk. Investigating Verbal Interaction in English.* London: Macmillan.

Larson, Mildred L.; Dodds, Lois. 1985. *Treasure in clay pots. An Amazon People on the Wheel of Change.* Palm Desert, CA & Dallas, TX: Person to Person Books.

Leet-Pellegrini, Helena M. 1980. "Conversational dominance as a function of gender and expertise". In: Howard Giles, W. Peter Robinson and Philip M. Smith (eds.): Language: Social Psychology Perspectives. New York, 97-104.

Lerner, Gene H. 1996. "On the "semi-permeable" character of grammatical units in conversation: conditional entry into the turn space of another speaker". In: Elinor Ochs, Emanuel A. Schegloff and Sandra A. Thompson: *Interaction and Grammar.* Cambridge: Cambridge University Press.

Levinson, Stephen. C. 1988. “Putting Linguistics on a proper footing: Exploration in Goffman’s concepts of participation”. In Paul Drew and A. Wootton (eds.): *Erving Goffman: Exploring the Interaction Order*, Cambridge: Polity Press, 161-227.

Makri-Tsilipakou, Marianthi. 1994. "Interruption revisited: Affiliative vs. disaffiliative intervention". Journal of Pragmatics 21 (1994), 401-426.

Marcoccia, Michel. 2004. "On-line Polylogues: conservation structure and participation framework in Internet Newsgroups". *Journal of Pragmatics* 36(1), 115-145.

Markee, Numa. 2000. *Conversation Analysis.* Mahwah, NJ: Erlbaum.

Mauranen, Anna and Elina Ranta (eds.) (2009): English as a lingua franca. Studies and Findings. Newcastle upon Tyne: Cambridge Scholars Publishing.

Meierkord, Christiane 2000. "Interpreting successful lingua franca interaction. An analysis of non-native/non-native small talk conversations in English". *Linguistik Online*, 5, 1/00.

Moerman, Michael 1987. *Talking Culture. Ethnography and conversation analysis*. Philadelphia: University of Pennsylvania Press.

Müller, Frank E., 1995. "Interaction et syntaxe. Structures de participation et structures syntaxiques dans la conversation á plusieurs participants". In: Véronique, D., Vion, R. (eds.), *Modéles de l'interaction verbale*. Publications de l'Université de Provence, Aix-en-Provence, 331-343.

Murray, Stephen; Covelli, Lucille. 1988. "Women and men speaking at the same time". In: *Journal of Pragmatics* 12, 103-111.

Oreström, Bengt. 1983. *Turn-taking in English Conversation*. Lund: CWK Gleerup.

Reisman, Karl. 1974. "Contrapuntal conversations in an Antiguan village". In: Richard Bauman and Joel Sherzer (eds.): *Explorations in the ethnography of speaking. Cambridge*: Cambrige University Press, 110-124.

Sacks, Harvey. 1984. "Notes on Methodology". In J. Maxwell Atkinson and John Heritage (eds.): *Structures of Social Action: Studies in conversation Analysis*. Cambridge: Cambridge University Press, 21-27.

Sacks, Harvey. 1967 [1995]. "March 2: Turn-taking; Collaborative utterances via appendor questions; Instructions; Directed utterances". In: Gail Jefferson (ed.): *Harvey Sacks. Lectures on Conversation*. Malden: Blackwell, 523-534.

Sacks, Harvey. 1964-72 [1995]. *Lectures on Conversation.* (ed. by Gail Jefferson). Malden: Blackwell.

Sacks, Harvey and Emanuel A. Schegloff. 1973. "Opening Up Closings". *Semiotica*, VIII, 4 (1973) 289-327.

Sacks, Harvey, Schegloff, Emanuel A.; Jefferson, Gail. 1974. "A simplest systematics for the organization of turn-taking for conversation". In: *Language* 50/4, 696-735.

Schegloff, Emanuel A. 1995. "Parties and Talking Together: Two ways in which numbers are significant for talk-in-interaction." In: Paul ten Have and George Psathas (eds.): *Situated Order*. Boston: University Press of America.

Schegloff, Emanuel A. 1996. "Issues of relevance for discourse analysis: Contingency in action, interaction and co-participant context." In. Hovy, Eduard H. and Donia R. Scott (eds.): *Computational and Conversational Discourse: Burning Issues? An Interdisciplinary Account.* Berlin: Springer, 3-35.

Schegloff, Emanuel A. 2000. "Overlapping talk and the organization of turn-taking for conversation." *Language in Society* 29, 1-63.

Schegloff, Emanuel A. 2002 [sic!]. "Accounts of Conduct in Interaction. Interruption, Overlap and Turn-Taking". In Jonathan H. Turner (ed.). *Handbook of Sociological Theory*. New York: Kluwer Academic/Plenum Publishers, 287-321.

Schegloff, Emanuel A.; Koshik, Irene; Jacoby, Sally; Olsher, David. 2002. "Conversation Analysis and Applied Linguistics". *Annual Review of Applied Linguistics* 22, 3-31.

Seedhouse, Paul. 2004. *The interactional architecture of the language classroom: a conversation analysis perspective.* Malden, Mass.: Blackwell.

Seidlhofer, Barbara. 2011. *Understanding English as a lingua franca*. Oxford: Oxford University Press.

Tannen, Deborah. 1994. *Gender and Discourse.* Oxford: Oxford University Press.

Tannen, Deborah. [2005]. *Conversational Style. Analyzing Talk among Friends.* 2nd edn. Oxford: Oxford University Press.

ten Have, Paul and George Psathas (eds.): *Situated Order. Studies in the Social Organization of Talk and Embodied Activities*. Boston: University Press of America.

Watts, Richard J. 1991. *Power in Family Discourse*. Berlin and New York: Mouton de Gruyter.

Wolfartsberger, Anita. 2011a. "ELF Business/Business ELF: Form and Function in Simultaneous Speech". In: Archibald, Alasdair; Cogo, Alessiona; Jenkins, Jennifer (eds.): *Latest Trends in ELF Research*, 163-183.

Wolfartsberger, Anita. 2011b. "Studying Turn-Taking in ELF: Raising the Issues". The 4th International Conference of English as a Lingua Franca (ELF4), Hong Kong, China, May 26-28.

Wolfartsberger, Anita. In prep. *Managing Meetings in ELF*. PhD Thesis, University of Vienna.

Wong, Jean; Olsher, David. 2000. "Reflections on Conversation Analysis and Nonnative Speaker Talk. An Interview with Emanuel A. Schegloff". *Issues in Applied Linguistics* 11(1), 111-128.

Wooffitt, Robin. 2005. *Conversation Analysis and Discourse Analysis. A Comparative and Critical Introduction*. London: Sage.

Zimmerman, Don H. 1988. On conversation: The conversation analytic perspective. Communication Yearbook II, 406-432. Newbury Park, CA: Sage.

Zimmermann, Don; Boden, Deirdre. 1991. "Structure in action: An introduction". In: Zimmermann, Don; Boden, Deirdre (eds.). *Talk and Social Structure. Studies in Ethnomethodology and Conversation Analysis.* Cambridge: Polity Press, 3-22.

A processing view on order in reversible and irreversible binomials

*Arne Lohmann, Vienna**

1. Introduction[1]

Binomials are a subclass of coordinate constructions, viz. the coordination of two single words which belong to the same form class; examples would be *hard and fast*, or *salt and pepper*. It has been a very popular research topic to investigate which factors determine the order of elements in irreversible instances within this class, such as *law and order*, or *nickel and dime* (e.g. Abraham 1950, Cooper & Ross 1975, Fenk-Oczlon 1989). While researchers noticed early that binomials vary with regard to their (ir-)reversibility (cf. Malkiel 1959), the question of how reversibility and the influences of ordering constraints interact has only recently been addressed (Lohmann 2011, Mollin 2012). The present paper sets out to contribute to this issue by exploring the similarities and differences between irreversible and reversible binomials with regard to ordering and to explain them from a processing perspective. This is done by analyzing corpus samples of both groups, remedying the shortcoming that in previous research reversible cases were neglected, which precluded an explicit comparison of both classes.[2] The

* The author's e-mail for correspondence: arne.lohmann@univie.ac.at

1 This paper is based on my PhD thesis, see Lohmann (2011). I am very grateful to the members of my dissertation committee, Thomas Berg, Britta Mondorf, Klaus-Uwe Panther, Günter Radden, and Tanja Kupisch for helpful comments and discussions.

2 It needs to be pointed out that Mollin (2012) also compares reversible and irreversible binomials. While there is thus a certain degree of overlap with the present study, there are differences pertaining both to theoretical perspective and to the empirical approach. The most important empirical difference is that Mollin's sample of binomials comprises solely binomials of considerable frequency, while the present study aims at comparing reversible ad hoc binomials of low-frequency, for which an online ordering process can be assumed, to frequent, irreversible ones. See also Note 6.

argument to be put forth is that ordering in both classes can be explained via properties of the processing system, with irreversibles representing 'fossilized processing preferences'. Let us first discuss the differences between both classes. Compare (1-3) which exemplify both types of binomials.

(1) house and home
(2) bed and board
(3) ...we can take over two of their sponsored events er which is golf and tennis and it would be something like ... (BNC, File FUG)

Examples (1-2) instantiate irreversible binomials whereas example (3) represents an instance of ad hoc coordination in speech and may certainly be reversed.

While the two classes are identical with regard to their syntactic structure, a number of characteristics which distinguish irreversible binomials from their reversible counterparts have been mentioned in previous research: The obvious distinction between the two classes is that irreversibles occur in only one order, or exhibit at least a very strong tendency to do so. A further property which contributes to this class's invariability of form is that the individual elements cannot be modified (compare He was willing to risk life and limb / *dear life and precious limb, example from Olsen 2002: 183). This formal conventionalization comes with a considerably high token frequency, certainly higher than the token frequency of a spontaneous ad hoc coordination, such as (3) (cf. Lambrecht 1984, Norrick 1988). With regard to semantics, it has been noticed that the meaning of many irreversible binomials is non-compositional, as e.g. house and home, or bed and board, respectively, do not just denote the sums of their respective constituents. Concluding, prototypical irreversible binomials are characterized by an invariable form and non-compositional semantics, rendering them similar to the class of idioms (see Lambrecht 1984, Norrick 1988, Masini 2006). It needs to be mentioned, however, that not all irreversible binomials necessarily fulfill all of the aforementioned characteristics. For instance, while the meaning of house and home is clearly idiomatic, the meaning of law and order is still fairly compositional. The notion of irreversible binomial is thus to be understood as a prototypical category.

Adopting a processing perspective reveals a further difference between the two classes, which makes their comparison particularly interesting. With reversible ad hoc constructions we can assume that the speaker performs an ordering process when producing the binomial. With this type of construction it is therefore possible to investigate the factors which underlie serialization of

elements during on-line processing. This characteristic distinguishes them from irreversible binomials: Since the latter group represents frequent, conventionalized and idiomatic instances, it can be assumed that they are stored as units in the mental lexicon (see Kuiper et al. 2007). Speakers thus 'reach for them' during production, but no longer perform an ordering process. This crucial distinction raises two interesting issues, which I will address in this paper.

First of all, regarding the strong focus on irreversible binomials in prior research, it is of imminent interest to find out which constraints influence the ordering in reversible ad hoc coordination. This means addressing the question of which factors impact the speaker during on-line serialization of elements in coordination contexts. General mechanisms of serialization during speech production are relevant for answering this question, which are to be detailed below.[3]

Second, since irreversible binomials are not the product of an on-line ordering decision, but represent lexicalized units, the question arises whether processing influences shown to influence ad hoc coordination are still observable in this class. Furthermore, these binomials may exhibit properties which render them particularly suitable for developing into lexicalized units. Two hypotheses on such properties are discussed in this paper. The first is that irreversibles are formed in analogy to monomorphemic words, which I term the 'lexical unit hypothesis'. The second hypothesis states that the very same constraints that underlie serialization in ad hoc coordination work as 'selection pressures' for irreversibles, yet their effects are much more pronounced in the latter class. The empirical analysis yields only little evidence for the first hypothesis, but largely supports the second one.With regard to an explanation of the empirical results, I argue that the found 'selection pressures' on irreversible binomials can be explained via preferences of the language production system.

The structure of the paper is as follows. Section 2 lists the influences on the order of constituents which will be empirically tested. In section 3 hypotheses on the differences between both groups are discussed. Section 4 elaborates on the data on which the empirical, corpus-linguistic investigation is based and explains the method employed during its analysis. Section 5 reports the results of the empirical analysis. Section 6 discusses the

[3] With regard to the methodology it is moreover important to note that any accurate description of the properties of irreversibles must use the class of reversibles as a backdrop to distinguish between the distinct characteristics of irreversibles and those of binomials as a whole.

differences between both groups, which are then explained from a processing perspective in section 7. Section 8 concludes the paper.

2. Influences on order in binomials

Before we embark on the empirical analysis it is necessary to provide a brief explanation of the ordering constraints to be tested by surveying previous research. The question of ordering of constituents in binomials is situated at the interface of research in linguistics and psycholinguistics. While linguists have focused on the properties of (mostly irreversible) binomials, general issues of serialization as discussed in psycholinguistics are obviously also relevant. Research in both areas is vast and therefore impossible to review in detail at this point (but see Lohmann 2011: Chapter 2 for an overview). In the following, I will therefore provide only a concise list of ordering constraints which have been discussed in previous works and which will be tested in the current empirical study (see also Benor & Levy 2006 for a detailed discussion of ordering constraints). If not obvious, I will also explain how these constraints were operationalized in coding the data.

Iconic sequencing: A certain order in extra-linguistic reality has been found to be reflected in the order of constituents (see Malkiel 1959, Benor & Levy 2006). This refers mostly to temporal order, e.g. *birth and death*, but may also refer to other scales, e.g. *eighth and ninth*.

Conceptual accessibility: The constituent which denotes the more accessible concept precedes the constituent denoting a concept of lesser accessibility (see Bock & Warren 1985). The contrasts considered here are: animate before inanimate, positive before negative, concrete before abstract, vertical before horizontal, prototype first, basic level before subordinate/superordinate level, proximal before distal, own before other, present generation before other (cf. Cooper & Ross 1975, Benor & Levy 2006).

Extralinguistic hierarchy: If one of the constituents' referents is ranked higher in an extra-linguistic hierarchy it precedes the other one. This variable refers to a male-first-bias, as well as to other hierarchies, e.g., *men and women*, *president and vice-president* (Sambur 1999, Benor & Levy 2006).

Short before long: The shorter constituent precedes the longer one (in number of syllables) (see Malkiel 1959, Cooper & Ross 1975), e.g. *law and order*.

Rhythm: The constituents are preferably ordered such that stressed and unstressed syllables alternate (McDonald et al. 1993, Benor & Levy 2006), e.g. salt and pepper vs. pepper and salt.
Avoidance of ultimate stress: The constituents are ordered such that a stressed ultimate syllable of the second constituent is avoided (Bolinger 1962), e.g. intents and purposes vs. purposes and intents.

Syllable Weight: The second constituent's main syllable is heavier than the first constituent's, as the second element is preferably stressed and heavy syllables attract stress (heaviness-to-stress principle) (Benor & Levy 2006). I considered syllables with long vowels (VV), a filled coda position (VC), or both (VVC) heavy syllables, while I considered syllables with a short vowel (V) followed by an ambisyllabic consonant light syllables, e.g. mother and child.

Number of initial consonants: The second element contains more initial consonants than the first (Cooper & Ross 1975), e.g. sea and ski.

Vowel length: The constituent with the longer main stressed vowel follows the one with the shorter vowel (Cooper & Ross 1975), e.g. stress and strain. The vowels were classified in the present study as follows: Short vowels: æ, e, ɪ, ʌ, ʊ; Long vowels: ɑ, e, i, o, u, ɔ, ɜ (see Benor & Levy 2006: 245).

Voicing of final consonant: As voiced consonants lengthen a preceding nucleus and voiceless consonants shorten it (cf. Peterson & Lehiste 1960), the constituents should be ordered such that the second constituent ends in a voiced consonant and the first constituent in a voiceless consonant (see Ross 1982, Lohmann 2011: 49), e.g. push and pull.

Vowel height/backness: The main vowel of the second constituent is lower and/or further back than the first constituent's main vowel (Cooper & Ross 1975: 71), e.g. dribs and drabs. Vowel height and backness were measured as the first and second formant frequency of the constituent's main vowel, respectively. Formant frequencies were taken from Steinlen (2002).

Sonority of initial consonant: The constituent with the more obstruent initial consonant follows the one with a less obstruent (more sonorous) beginning (Cooper & Ross 1975), e.g. wheel and deal. The following sonority scale was applied, from most sonorous to most obstruent: h > j > w > r > l > nasals > fricatives > stops (cf. Benor & Levy 2006: 250).

Sonority of final consonant: The second constituent ends in a more sonorous consonant (Cooper & Ross 1975, Wright et al. 2005), e.g. safe and sane.
Frequency: Constituents with a higher frequency precede those with a lower token frequency (Fenk-Oczlon 1989). The frequency of every constituent was measured as token frequency of the exact word form in the BNC.

3. Deriving hypotheses on differences between reversible and irreversible binomials

Since the main aim of this article is to compare reversible and irreversible binomials with regard to ordering constraints, let us turn to possible hypotheses on differences and similarities between both groups. To the best of my knowledge no explicit suggestions have been made in the literature; there are, however, certain assumptions mentioned in previous research from which two testable hypotheses can be derived.

The first hypothesis takes as its starting point the assumption that the formation of binomials may be explained by properties of monomorphemic words, as some of the phonological properties of the latter class may also be found in the former (see Müller 1997, Wright et al. 2005). This possibility also suggests a hypothesis on the differences between reversibles and irreversibles, viz. that irreversibles exhibit more similarity to monomorphemic words, as they are more strongly lexicalized.

The logic underlying this claim is as follows: Since irreversible binomials are characterized by an invariable form and often non-compositional semantics, it can be assumed that they are stored as units in the mental lexicon, similar to words (see above, see also Müller 1997: 19-21). As irreversible binomials become part of the lexicon, they inherit phonological properties of other units in the lexicon and thus are formed after the 'model' of monomorphemic words. Hence, as they are stored like words, also their form becomes more 'word-like', by virtue of analogy. In contrast, reversible binomials should not be as strongly influenced by this process, as they do not represent lexicalized units. We may term this assumption on possible differences between both groups the 'lexical unit hypothesis' (LUH).

In order to delimit the scope of this hypothesis I rely on Wright et al. (2005: 536), who conducted a study on the properties of English monomorphemic words based on the CELEX database, from which they then derived ordering tendencies. Translating these into assumptions on differences between irreversibles and reversibles leads to the following sub-hypotheses: Since English monomorphemic words are characterized by initial

consonant clusters and obstruent phonemes, the two corresponding ordering constraints should yield a stronger influence on irreversible binomials (see above Section 2). Irreversible binomials are furthermore predicted to prefer a sonorous final segment, which is also a property of monomorphemic words.[4] In conclusion, the prediction of LUH is that the three aforementioned ordering constraints yield more pronounced effects in the sample of irreversibles.

The second hypothesis I will test is inspired by Pinker & Birdsong (1979), who find that the ordering of nonce words in coordination is sensitive to many of the ordering constraints I mentioned above (see Section 2). Adopting a somewhat Darwinian perspective, they term these constraints 'selection pressures' which weed out some and facilitate other orderings, as these are easier to process (Pinker & Birdsong 1979: 506-7). An extension of this hypothesis in the present context is that these selection pressures are adhered to more strictly in the group of irreversibles. The logic would be as follows: Certain orderings in ad hoc coordination are more preferable for the language user than others, namely those which adhere to ordering constraints. Some of these preferred instances become conventionalized and irreversible, concomitant with a high frequency of use. It seems only logical that the linguistic community would choose those instances for this development which are easiest to produce and process – in conforming best to existing constraints. The prediction of what I term the 'selection pressures hypothesis' would thus be that both groups adhere largely to the same ordering constraints, yet their effects are much more pronounced in the group of irreversibles.

In the following I will present an empirical analysis of ordering constraints in both reversible and irreversible binomials and discuss whether the predictions of the two hypotheses are borne out. Note that the 'lexical unit hypothesis' and the 'selection pressures hypothesis' are not mutually exclusive but may complement each other. It is even possible to integrate the first into the latter, as a similarity to monomorphemic words may constitute one selection pressure in the sense outlined above.

4 A further analogy may be found in the stress pattern of irreversibles. Müller (1997) argues German binomials to exhibit the same stress pattern as equally long polysyllabic, but monomorphemic words. The standard of comparison in our case would be monomorphemic words which are four to five syllables long, as the majority of irreversibles coordinate a monosyllabic and a disyllabic constituent, or two disyllabic constituents. What renders a comparison problematic is the fact that longer polysyllabic words in English do not exhibit a consistent stress pattern and monomorphemic words of these lengths are infrequent. An explanation of the stress pattern of irreversibles in terms of the LUH is thus not very plausible for English, as no systematic pattern which would be frequent enough to serve as a model exists.

4. Data and method

The present analysis of ordering in binomials employs corpus-linguistic methods, thus is based on naturalistic usage data. The choice fell on the British National Corpus, as it is large enough to yield a fair amount of irreversible binomials and is evenly balanced across different genres. In order to create samples of irreversibles and reversibles, respectively, different parts of the corpus were employed. With reversible ad hoc coordination my aim is to explore the influences that underlie serialization during on-line processing. Therefore spoken data is the medium of choice, as it keeps possible editing influences to a minimum. Reversible binomials were consequently sampled from the spoken part of the BNC. For the creation of a sample of irreversible binomials the entire corpus was used, however, as their identification required a large corpus. Since homogeneous samples were aimed at, the analysis is restricted to nominal binomials coordinated by and (N and N) which make up the by far largest group among binomials as a whole (cf. Mollin 2012).

In order to compare the two groups, an empirical method to identify and extract only the irreversibles from the corpus is needed. This issue has not been addressed yet in prior research, as irreversibles were identified introspectively. However, such approaches to reversibility rest on highly subjective assessments and are therefore prone to error. In tackling this problem by using corpus data, the most obvious empirical approach would be to test whether a given binomial occurs in only one order in a large corpus, thus is practically irreversible. This seemingly attractive solution entails two problems, however, and therefore needs to be modified:

The first is that, if applied strictly, types which are reversed only very rarely, possibly only once in the corpus, would be excluded. Hence even a rare reversal for rhetorical reasons would remove the data point from our sample of irreversibles. A famous example of such a reversal is Samuel Beckett's collection of dramas entitled ends and odds, a word play on the irreversible binomial odds and ends which certainly does not render the latter reversible.[5] Therefore I will consider those types irreversible for which one ordering makes up ninety percent of its hits in the corpus data. This operationalization leaves some room for exceptional reversals while still capturing those binomials with a strong ordering bias.

The second problem that needs to be addressed is that of low-frequency types. If we just focused on reversibility without considering frequency, misleading results would be obtained for instances of low token frequency.

5 I thank Britta Mondorf for bringing this example to my attention.

For example, the coordination viola and harp occurs three times in the BNC but never in reverse order. A reversal is, however, certainly possible; chances are high that it is simply not found in the corpus data due to chance, as it represents an infrequent type containing two lexemes that are rarely combined. Only if a certain frequency threshold is surpassed, can we be sure that the corpus finding of irreversibility is not due to chance. Implementing such a threshold furthermore captures another defining characteristic of irreversible biniomials, viz. their conventionalization and a concomitant high frequency of use (see above). For these reasons the second empirical criterion I apply to identify irreversibles is that a frequency threshold of 10 per 100 million words has to be exceeded.

The corpus extraction procedure was thus as follows: All coordinations of two nouns linked by and which instantiate individual noun phrases and form a superordinate NP which does not contain additional material were extracted from the BNC.[6] The two criteria for the extraction of irreversibles were applied; 259 types fulfilled both of them and were therefore kept in the sample of irreversible binomials. The example heaven and earth may illustrate the identification process. The ordering heaven and earth occurs 66 times in the BNC, while the reversal earth and heaven is instantiated 3 times. While the binomial is thus not irreversible in a strict sense, one ordering makes up 95.7%, it thus exhibits a very strong ordering bias. Since heaven and earth also occurs frequently enough to surpass the frequency threshold, it qualifies as an irreversible binomial applying the suggested operationalization.

For the sample of reversibles, similarly all N and N instances were extracted, this time from the spoken part of the BNC, applying the same criteria to weed out false hits. All irreversible types identified in the first step of the analysis were removed from the sample of reversibles. Furthermore, all types with a token frequency of higher than 10 per 100 million, whether reversible or not, were not considered in this sample, as with these lexical unit status cannot be ruled out: Even if a certain construction does not exhibit a strong tendency towards one of two possible orderings, it is still conceivable that a language user has both orderings stored as units in the mental lexicon. Such effects are however unwanted in the sample of reversible ad hoc orderings. After having removed these binomials, every other hit was kept for

6 This means that coordinations of more than two elements, e.g. trinomials, were not considered. Furthermore proper names, e.g. Guns and Roses were weeded out. Also binomials containing extender phrases such as and things or and stuff and of course also repetitions such as apple and apple were not considered.

further analysis, resulting in a sample of 1,109 reversible types.[7] It needs to be pointed out that both samples are type and not token samples, where one type is one particular ordering, e.g. heaven and earth. Using token samples would mean that certain extraordinarily frequent binomials in the sample of irreversibles (e.g. law and order, which occurs 598 times in the BNC) may distort the results, as it is possible that individual binomials are influenced by idiosyncratic constraints. Since the main thrust of this paper is to find out about the general characteristics of the group of irreversibles as a whole, however, type samples are more adequate.

The two samples were coded for the ordering constraints mentioned above. Details pertaining to the operationalization of individual constraints are given in the table below. Note that all ordering constraints are based on possible contrasts between the two constituents, e.g. one constituent denotes the more accessible concept, or is longer or more frequent than the other. Coding for the relevant variables therefore means coding possible differences on the respective dimensions expressed by the ordering constraints. The result of the coding process is thus a vector of differences between the elements of the respective binomials.

Let me exemplify the coding process, starting with the categorical variable Conceptual accessibility. If, for example, the first constituent denotes an animate referent and the second an inanimate referent, the conceptual accessibility criterion was adhered to and coded by the value (1); if the reverse order was encountered and the constraint was thus violated, it received the coding (-1). If the criterion did not apply, as no difference in conceptual accessibility between the two constituents was observed, it was coded (0). The same procedure was applied for all categorical variables, i.e. those constraints which are either categorically adhered to or violated, but do not allow for more fine-grained scalar distinctions. The procedure is a little

7 Although the present operationalization results in a cut-off point which divides linguistic examples into two categories, I do not wish to propagate a binary view on formulaicity or lexicalization. On the contrary, as has been shown for other fixed expressions, we are most likely dealing with a continuum of free and fixed coordinations (see Wulff 2008). Still, in order to distinguish between the two groups, for which an (at least gradually) different storage and therefore processing is likely, some kind of operationalization is necessary. However, I am the first to admit that the one suggested here is no more than a heuristic approach which does not necessarily mirror cognitive and psychological reality adequately. The present analysis differs from Mollin's (2012) approach who employed four reversibility categories. While her approach is thus more fine-grained than the present one, the exact calculation of reversibility requires binomials which are considerably frequent. Since frequent binomials may be stored as units in the mental lexicon (conceivably even both orderings), this approach would run counter to the aim of exploring order in spontaneous ad hoc coordination, for which we may assume an online serialization process.

different with scalar variables, e.g. the short-before-long constraint. In these cases first the length values for both constituents were coded and in a second step the difference in length between both constituents was calculated. For example, with the binomial salt and pepper, the lengths of the constituents in syllables are (1), and (2) respectively. Since the relevant ordering constraint predicts the shorter constituent to precede the longer one, the value of the first constituent was subtracted from the second, yielding a value of (1) in the example. This procedure ensures that positive values denote an adherence to the short-before-long constraint and negative values a violation of it. The following table illustrates the length coding of two binomials. Other scalar variables were treated similarly.

Binomial	Ordering constraint	Coding
Salt and pepper	Short>long	+1
Margarine and salt	Short>long	-2

Table 1. Sample coding of the length constraint

Having explained the general coding procedure, information on the level of measurement and the possible difference values for every individual variable is given in the table 2 below.

Every individual binomial was coded for the fifteen ordering constraints tested. This coding procedure resulted in a vector of (difference) values (one for each binomial type analyzed) for each ordering constraint, which consists of the possible values given in the rightmost column above. For all constraints positive values denote an adherence to the ordering constraint, while negative values mean a violation of it. For example the Short>long vector comprises all differences in length between the binomials, ranging from x to y. If more binomials exhibit a short-before-long than a long-before-short pattern, positive values will outweigh negative ones.

The fifteen vectors were then entered into a multifactorial logistic regression model without intercept. Ordinary logistic regression models, i.e. models with an intercept, are a fairly established method to model linguistic choices with a binary outcome (see Baayen 2008, for an introduction). The interceptless model represents a variant of this method which is particularly apt to deal with ordering problems (see Levy in progress Ch. 6.8.4 and Lohmann 2011 for a detailed description, see also Wiechmann & Lohmann

Variable	Operationalization	Level	Values
Iconic Sequencing	This constraint was coded 'adhered to', if the order of element mirrored the order in extra lingui tic reality or 'violated', if the order was reversed.	categorical	1, 0, -1
Conceptual Accessibility	The aforementioned contrasts were applied and coded whether adhered to, violated or inapplicable.	categorical	1, 0, -1
Hierarchy	A sequence of higher to lower rank meant that the constraint was 'adhered to', the opposite order instantiated a violation.	categorical	1, 0, -1
Rhythm	If the actualized order instantiated an alternation of stressed and unstressed syllables, but the alternative order would not, this constraint was coded 'adhered to'. The opposite case was coded a violation.	categorical	1, 0, -1
Syllable weight	A light-to-heavy sequence was coded as adherence to the constraint and a reversal as a violation.	categorical	1, 0, -1
Avoidance of ultimate stress	If the constituents were ordered such that ultimate stress was avoided, this variable was coded 'adhered to'. In the opposite case it was coded 'violated'.	categorical	1, 0, -1
Short>long	Length in number of syllables was coded for every constituent. A's length was subtracted from B's (B-A).	scalar	-5 to +4
Sonorous initial consonant	Values from 1 (most obstruent) to 8 (most sonorous) were assigned to the individual consonants applying the sonority scale provided above. A's value was subtracted from B's (B-A).	scalar	-7 to +7
Sonorous final consonant	Similarly, obstruency values were assigned to final consonants. The difference (B-A) was calculated.	scalar	-7 to +7
Voicing of final consonant	The following sequences were considered as adherence to the constraint: voiced/unvoiced; vowel/voiced; unvoiced/vowel; their opposites as violations.	categorical	1, 0, -1
Initial consonants	The difference in number of initial consonants was calculated (B-A).	scalar	-2 to +3
Vowel length	A sequence of short before long main vowel was coded 'adhered to', the reverse order was coded 'violated'.	categorical	1, 0, -1
Vowel backness	F2 values of both constituents' main vowels were divided by 100 and the difference (A-B) was calculated.	scalar	-14.7 to +12.5
Vowel height	F1 values of both constituents' main vowels were divided by 100 and the difference (B-A) calculated.	scalar	-4.8 to +4.5
Frequency	B's log-transformed token frequency was subtracted from A's (A-B).	scalar	-3.2 to +3.7

Table 2. Overview of tested ordering constraints

forthcoming for another application of the method).[8] Models were built for the sample of reversible and irreversible binomials respectively, which calculate the statistical significance of the individual ordering constraints. A statistically significant result of a constraint means that it helps us to predict the ordering of binomials correctly in the respective sample. In fitting the model, statistically non-significant variables were removed using stepwise backwards elimination, until minimal adequate models were arrived at, i.e. models which contain solely significant predictors (cf. Crawley 2005).

5. Results

The elimination of non-significant variables resulted in two minimal adequate models, one for each sample, which are reported in the table below. Even a quick glance at the tables above reveals that not all variables made it into the minimal adequate models, as a number of variables yielded non-significant results for reversibles and irreversibles and were therefore eliminated. These are Number of initial consonants, Avoidance of ultimate stress, Vowel length, Sonority of initial consonant, Vowel height/backness. This result means that no evidence for an influence of these variables on ordering was found on the basis of the two samples.[9] With regard to the remaining variables, a great overlap can be observed between the two models, as all variables which yield a significant result in the sample of ad hoc binomials are also significant in the sample of irreversibles. These shared variables comprise all semantic variables, i.e. Iconic sequencing, Conceptual accessibility, and Extralinguistic hierarchy. Also Syllable weight, the Short-before-long bias and Frequency significantly influence reversible as well as irreversible binomials. Two variables remain only in the model for irreversibles, viz. Rhythm and Sonority of final consonant.

8 The present method represents a refinement of the method applied in Benor & Levy (2006) in that it takes into account categorical and also scalar variables.

9 However, it needs to be pointed out that this non-significant result does not rule out an influence of these ordering constraints in the population of binomials. It merely means that no evidence was obtained on the basis of the corpus samples I used. It remains possible that effects of the mentioned ordering constraints are found in an empirical study with greater power (in a statistical sense). Since power is primarily influenced by sample size, an empirical study based on larger samples may further explore the workings of these constraints.

	Irreversible binomials			Reversible ad hoc binomials		
Variable	Coefficient	Odds ratio	p	Coefficient	Odds ratio	p
Iconic Sequencing	3.13	22.8	**	1.46	4.32	**
Conceptual Accessibility	1.69	5.43	**	0.45	1.56	*
Hierarchy	1.92	6.8	***	0.74	2.10	**
Rhythm	0.97	2.65	*	-	-	-
Syllable Weight	1.73	5.66	***	0.27	1.31	+
Short>long	1.02	2.78	***	0.16	1.18	*
Sonorous final consonant	0.37	1.45	*	-	-	-
Frequency	0.74	2.09	*	0.12	1.12	+
N	259			1109		
% correct	83.8			60.5		
*** p<0.001	** p<0.01	*p <0.05		+p<0.1	- p>0.1	

Table 3. Minimal adequate models for the samples of irreversible and reversible binomials

There are a number of model values given in the table which require more explanation. One is the ratio of correctly predicted orderings. The aim of any statistical model is to accurately predict the values of the dependent variable, which in the present case is the ordering of elements observed in a given binomial. The ratio of correctly predicted orderings informs us which percentage of the orderings the model can correctly predict on the basis of the constraints which feature in the respective models. This value of predictive accuracy is conspicuously higher for the sample of irreversibles, in allowing us to correctly predict 83.8% of the data. This means that if the model is given the values for Frequency, Short-before-long and all other variables in the model, it can predict the ordering of elements, and these predictions would be correct for 83.8% of all binomials in the sample. For reversible binomials we

obtain only a value of 60.5%.[10] Both models, however, represent significant improvements over the baseline value of 50% which would be obtained by simply guessing the order of constituents in the samples.

Other important values are the coefficient values and odds ratios for the different variables. Both are measures of effect size and thus inform us about the strength of the respective ordering constraints. Coefficient values run from $-\infty$ to $+\infty$, with more extreme values indicating stronger effects. In the present context positive values indicate an effect in the predicted direction, e.g. short-before-long, while negative values would denote an effect in the opposite direction, e.g. long-before-short. Since all coefficient values are >0, it follows that all variables influence ordering in the predicted direction. High positive values mean that the respective constraint is only rarely violated, while values closer to (0) signify a higher ratio of violation. Odds ratios are to be interpreted a little differently. They indicate how the odds for a certain outcome change through the influence of an independent variable. Odds greater than (1) indicate an effect in the predicted direction; the larger the value is, the stronger the effect. Values below (0) would indicate violations of ordering constraints. Corresponding to the coefficient values, all odds ratios are >(1). Comparing these two measures of effect size across the two models yields the interesting result that their values are uniformly higher in the sample for irreversibles, which means that the ordering constraints yield stronger effects with irreversibles, compared to reversible binomials (similar to results obtained by Mollin 2012). In the following, I will further interpret these differences and discuss what the results mean for the hypotheses on possible differences between the two classes.

6. Discussion

Let us at this point discuss what the results obtained mean for the questions I asked in the introduction on the differences between irreversibles and reversibles and, in particular, how the two hypotheses on these differences introduced in Section 3 fare against the data.

The first question posed was whether the ordering constraints underlying ordering in reversible binomials also determine order in irreversible cases. Since most of the variables are shared between the two models reported above, this question can largely be answered in the affirmative. This result is

[10] Interestingly, Benor & Levy's (2006) model, which jointly considered both groups, made around 77% correct predictions, a predictive accuracy which is in between the two values.

of great relevance for the processing perspective I adopt: Since the ordering of constituents in reversible ad hoc coordination can be assumed to be an on-line operation during language production, the results for that sample inform us about the forces at work during that serialization process. The fact that irreversible instances, for which an on-line ordering is no longer necessary, are affected largely by the same constraints indicates that these processing influences are also relevant for the emergence of fossilized, irreversible binomials. I will discuss this issue in detail in the next section.

The second question pertains to possible differences between the two classes and how these may be explained. Recall that I suggested two possible hypotheses on these differences, the 'lexical unit hypothesis' and the 'selection pressures hypothesis'. The first hypothesis hinges on the argument that irreversibles resemble monomorphemic words to a greater extent than reversible cases. The results for the variable Sonority of final consonant is perfectly in agreement with LUH: Similar to English monomorphemic words, irreversible binomials prefer a sonorous final segment, while such a trend is not observed in reversibles. Other hypothesized properties were, however, not found: A tendency for initial consonant clusters or for obstruent beginnings was found neither with irreversibles nor with reversibles. Overall, there seems to be only little evidence for the hypothesis that irreversibles exhibit a stronger resemblance to monomorphemic words.

The second hypothesis assumes selection pressures to be at work which underlie the emergence of irreversible binomials. This hypothesis states that the same factors underlie ordering in both groups, yet their influence should be more pronounced in irreversibles. While we already observed that the first part of this argument is borne out by the results, let us have a closer look at the second part. There are two possible ways in which one could interpret the assumption of a greater effect of ordering constraints in irreversibles: The first is that the adherence rate of ordering constraints is higher in the group of irreversibles, i.e. we would expect to find fewer violations of ordering constraints. A second possibility is that the ordering constraints influence a greater share of data points in the sample of irreversibles, irrespective of adherence to them.

Let me explain this difference using the conceptual accessibility constraint as an example, which predicts the more accessible concept to precede the constituent which is less accessible. The first interpretation means that if there is such a contrast between to-be-ordered constituents it should be less often violated in the sample of irreversibles as compared to the reversibles sample. The second interpretation refers to a possibly different scope of the conceptual accessibility constraint. This means that more binomials in the

sample of irreversibles exhibit conceptual accessibility contrasts than in the sample of reversibles, i.e. the constraint would apply more often. I will discuss these assumptions in turn, starting with the question of violation/adherence to ordering constraints. This question can be answered by comparing the effect sizes of constraints across irreversible and reversible binomials. Effect sizes in the model output are given as coefficient values or odds ratios, which are a direct expression of violation/adherence to ordering constraints – the greater the effect size, the fewer violations of the respective ordering constraint. Averaged coefficients of shared predictors are displayed in the following figure.[11]

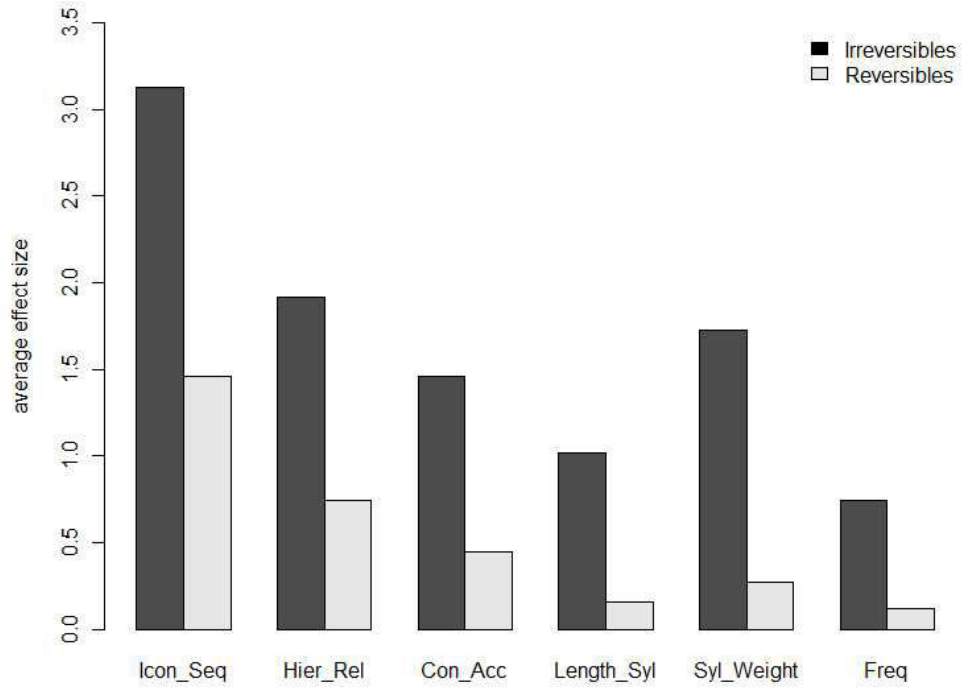


Figure 1. Average effect sizes of ordering constraints in irreversible and reversible binomials

The bars in the figure above are uniformly higher for irreversible binomials, which denotes greater effect sizes for all constraints in that class, indicating that these are much less often violated in irreversible binomials. This result thus confirms the selection pressures hypothesis: irreversible binomials adhere to ordering constraints more strictly than their reversible counterparts.

11 To allow also a comparison of the effects' strengths within the individual samples, all effect sizes were standardized by multiplying the respective coefficients by the average absolute value of the input vector.

Let us turn to the second possible interpretation of a greater pronouncedness of constraints, which states that irreversible binomials are more often affected by ordering constraints than reversibles. We may start this comparison with the semantic constraints. The following figure displays the percentages of data points which are influenced by the semantic ordering constraints we investigated in irreversibles and reversibles, regardless of whether these constraints are adhered to or not.

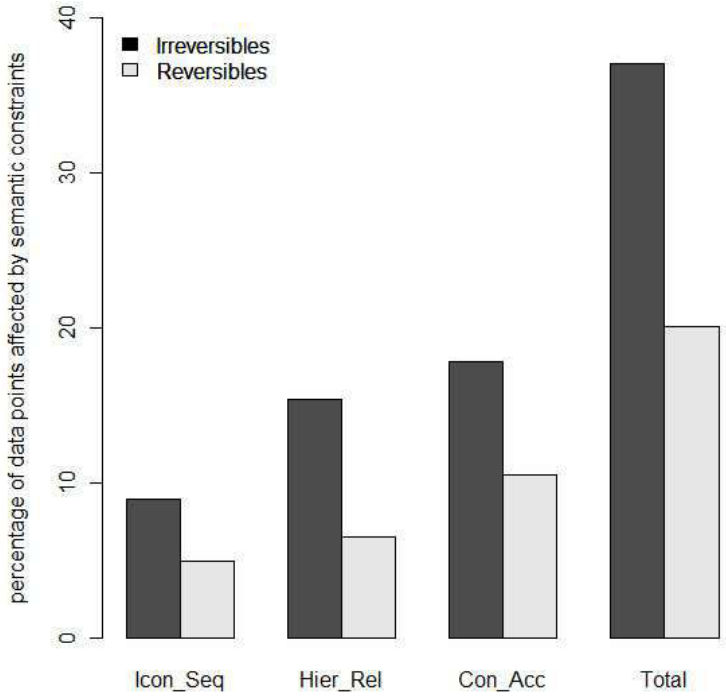


Figure 2. Semantic constraints in irreversibles and reversibles

Again the bars are higher for irreversibles, showing that all semantic factors are more frequently active in irreversibles than in reversibles.[12] The two rightmost bars indicate that in 37.1% of all irreversible binomials at least one semantic constraint applies, while this is true for only 20.1% of reversibles. Since all ordering constraints are motivated via certain differences/contrasts between the to-be-ordered elements (see list of constraints above), this finding means that elements in irreversibles exhibit greater differences on the relevant semantic dimensions than the constituents in reversibles. A second area we

12 Chi-square tests yield significant results for all pair-wise comparisons: Iconic Sequencing: χ^2=4.94, df=1, p=0.027, φ=0.07. Extralinguistic Hierarchy: χ^2=20.67, df=1, p<0.01, φ=0.12, Conceptual Accessibility: χ^2=9.67, df=1, p<0.01, φ=0.08, Total: χ^2=29.50, df=1, p<0.01 φ=0.15.

may examine is whether these greater contrasts hold also for the two scalar variables length and frequency. The following barplot displays the average differences between the two constituents along these dimensions for irreversible and reversible binomials.

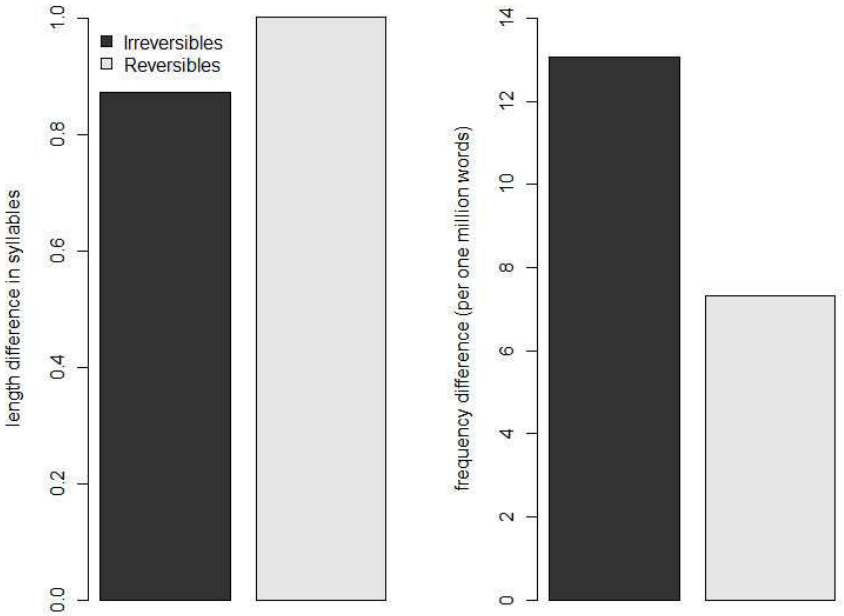


Figure 3. Length and frequency differences in irreversibles and reversibles

Regarding length, the average difference between constituents in irreversibles is 0.87 syllables, while it is 1.0 syllables in reversibles, in contrast to our assumption. This difference is, however, only marginally significant ($t_{two\text{-}tailed}$ = -1.77, df = 372, p = 0.08). With regard to frequency the assumption of greater contrasts in irreversibles is borne out, as the frequency difference is more pronounced in that group (mean difference = 13.06 per one million words), as compared to the group of reversibles (mean difference = 7.32 per one million words), which constitutes a statistically significant difference ($t_{two\text{-}tailed}$ = 2.38, df = 308, p = 0.02). In summary, almost all comparisons confirm the second assumption: The to-be-coordinated constituents in irreversibles exhibit greater contrasts or dissimilarities on all semantic dimensions, as well as with regard to their lexical frequency.

In conclusion, both possible interpretations of a greater effect of ordering constraints on irreversibles are borne out by the data: ordering constraints are much less often violated in the class of irreversibles as compared to reversible binomials. Furthermore, the constraints are more often active in the former group which means that the elements in irreversible binomials exhibit more, or greater contrasts than constituents of reversible

binomials. Due to these characteristics, the order of irreversible binomials can be predicted much more precisely than the order of reversibles, which is reflected in different values of predictive accuracy: While the model for irreversibles allows us to predict 83.8% of the cases correctly, the predictive accuracy for our sample of reversibles is considerably lower with 60.5%. These findings are in line with the predictions of the selection pressures hypothesis. Ordering constraints active during on-line ordering decisions seem to work as selection pressures for irreversible binomials: those binomials which best conform to them stand a greater chance of becoming irreversible.

7. A processing explanation

The preceding discussion of the results has shown that there is little evidence for the hypothesis that irreversible binomials are formed in analogy to monomorphemic words (Lexical Unit Hypothesis); however, the predictions of what I termed the 'selection pressures hypothesis' are borne out. The question that remains to be answered is how the finding of more pronounced effects of ordering constraints in irreversible binomials may be explained. In the following section I put forth the argument that the differences between the two groups ultimately stem from processing preferences, which underlie the emergence of irreversible binomials.

In order to flesh out this argument, I will first describe the processing of reversibles, before I turn to an explanation of the properties of irreversible binomials. Recall that I argued above that during the production of reversible ad hoc binomials, the processing system has to perform an ordering decision, i.e. choosing which constituent is produced first. It is a widespread view in language production research (cf. Bock 1982, Bock & Levelt 1994, Dell 1986, inter alia) that the ordering of constituents is contingent on their respective activation level:[13] Those constituents which are more highly activated at the time of production occupy early positions in a sentence, or generally in syntagmatic strings. Corresponding to that view I have shown elsewhere (Lohmann 2011) that the ordering constraints at work in reversible binomials can be related to the respective activation levels of the to-be-ordered constituents in a spreading activation model of language production (e.g. Dell 1986, Dell et al. 1997, Dell & O'Seaghdha 1994). Let me briefly

[13] The term activation is used in this paper in a wider sense, denoting both activation which is dependent on discourse context (e.g. through previous mention), as well as inherent activation due to inherent properties of the constituents, e.g. animacy.

describe the general logic of this argument. Remember that the constraints underlying ordering indicate contrasts between the two constituents. From the results of the quantitative analysis it can be inferred that in comparison to the second constituent the first one tends to be shorter, more frequent, conceptually more accessible, occupies a higher position in an extra-linguistic hierarchy, etc. These contrasts translate to activation differences between the two constituents, as for instance it is a well-established finding that both frequent words and/or short words are more easily retrieved from the mental lexicon and can thus be viewed to be more highly activated. The same holds for the conceptual differences subsumed under the variable Conceptual accessibility. Although requiring additional assumptions, the same argument can be made for the other relevant variables (see Lohmann 2011, Chapter 10, for a detailed discussion). The ordering constraints thus denote an activation difference between the two constituents. See the following figure for an illustration.

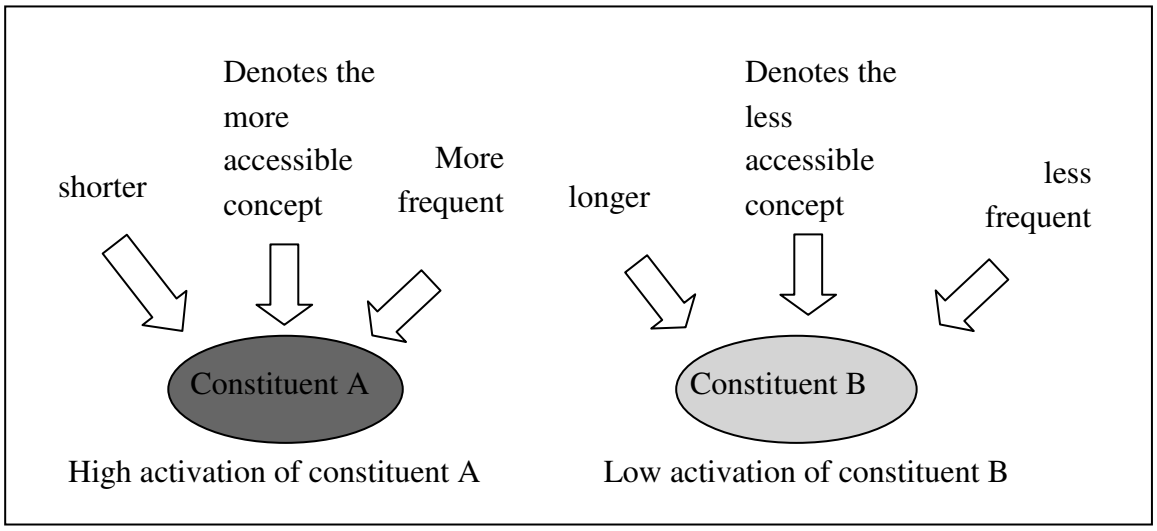


Figure 4. Activation differences of constituents in reversible binomials

Consequently, the ordering process in reversibles can be assumed to work as follows: During the production of reversible binomials, the processing system activates both constituents which thus compete for first mention. The constituent with a higher activation level, due to the characteristics denoted by the ordering constraints, is then selected and placed in early position. This account of the ordering process provides a basis for the explanation of the characteristics of irreversible binomials.

The first characteristic of irreversibles to be explained is that the shared ordering influences are more seldom violated in that class as compared to

reversible cases. In terms of activation this means that the constituent which is assumed to be more highly activated is more often found in first position in that class as compared to reversible binomials. This can be explained by situational influences on activation levels that are not captured by the ordering constraints and which influence the two classes differently: During ad hoc serialization in spontaneous speech there are confounding situational influences which impact the activation levels of the constituents and therefore result in a different than expected order of constituents. Such differences on the activation levels of the constituents may for example result from previous production processes, or the second constituent of the binomial could have been added 'on the fly' after the first was already produced. Irreversibles, however, can be assumed to not be influenced by these situational influences to the same degree, as they have undergone a lexicalization process. We may view this process as a collaborative production/processing effort of many production instances which ultimately result in a formulaic, irreversible unit. Although every individual instance of production during this process is similarly affected by situational influences, on the whole it is the order which best conforms to the high-low activation pattern as evidenced by the ordering constraints that is more frequently produced. Therefore this ordering stands a greater chance of 'fossilizing', i.e. developing into an irreversible, formulaic unit. Concluding, a mitigation of confounding situational factors in the course of the emergence of an irreversible binomial may explain the higher adherence rate to ordering constraints in that class.

The second difference between the two groups to be accounted for is that irreversibles are more often influenced by ordering constraints irrespective of their adherence. I argue that this characteristic can also be explained by the architecture of the processing system: Since ordering constraints can be related to activation differences between constituents and as these constraints affect irreversibles more often, I conclude that this class exhibits more pronounced activation differences between constituents. The argument to be put forth is that these greater differences mean less competition between the constituents for first position, which contributes to a smoother production process. Central to this argument is the notion of 'inhibition', an important architectural feature of interactive activation models of language production. Within this class of models, it is assumed that there are inhibitory links between forms on the same level, e.g. between two words of the same form class (cf. e.g. Dell & O'Seaghdha 1994). If a word (or other form) is activated for production, it inhibits the activation of its competitors via these links, which ensures that only one form is eventually selected for production.

These inhibitory links mean that during the production of a binomial its two constituents inhibit each other as they are activated for production. As one of them gains excitatory activation, it sends inhibitory activation to the other. Crucially, the extent and direction of inhibition is dependent on the activation differences between the two constituents. In a situation of large activation differences, which is the case if many ordering constraints apply, one constituent has a much higher activation level than its competitor and thus strongly inhibits the lesser activated constituent. In contrast, if both constituents have nearly equal activation levels, as few or no ordering constraints apply, also the inhibition of both constituents is nearly equal, i.e. there is stronger mutual inhibition between the two competitors. Both of these possibilities are illustrated in the following figure.

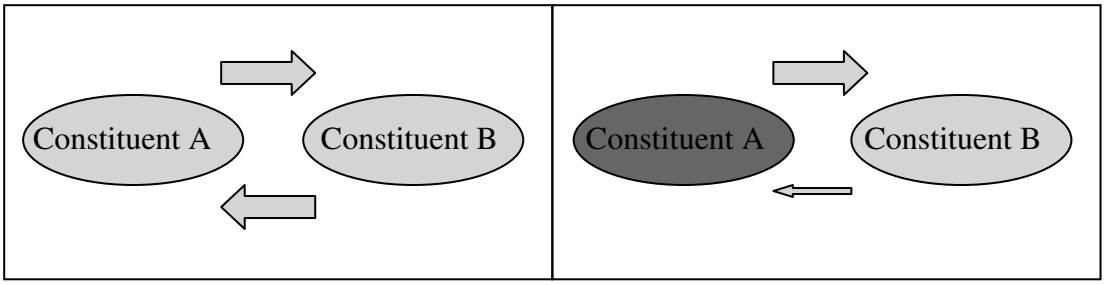


Figure 5. Activation differences and inhibition of constituents

The left part of the figure illustrates the situation of equal activation levels (inhibitory activation flow marked by arrows). Due to strong mutual inhibition it takes longer until one of the two constituents reaches a high enough activation level to be selected for first mention, which slows down the production of the binomial as a whole. Conversely, if there are large activation differences (illustrated by different shades in the right hand part of the figure above), only one constituent is strongly inhibited, and the selection of the constituent to be produced first may proceed largely unimpeded, resulting in an overall smooth production process. Consequently, as activation differences are much more pronounced in irreversible binomials, reflected in greater differences between constituents as denoted by the aforementioned ordering constraints, it follows that these are easier to produce and process.

In summary, both findings, the greater adherence rate in irreversibles and the greater differences between the constituents in that class can be related to mechanisms of the processing system, with irreversible binomials conforming better to its preferences. Hence it can be assumed that processing preferences are the 'true' selection pressure for the class of irreversibles, as it seems logical that the speech community would produce those instances more

frequently which are easier to process and thereby facilitate their development into formulaic constructions. Once this lexicalization process is completed, irreversible binomials can be considered units in the mental lexicon for which the speaker may reach, without performing an ordering process anymore. Nevertheless, many characteristics of this process can still be observed in them, which is why we may view irreversible binomials as representing fossilized processing preferences.

8. Conclusion

The present paper compared reversible and irreversible binomials with regard to the effects of ordering constraints. Two hypotheses on possible differences between the classes were tested: The 'lexical unit hypothesis', which states that irreversible binomials are similar to monomorphemic words, and the 'selection pressures hypothesis', which predicts that ordering constraints are shared between both classes, yet their effects should be more pronounced in irreversible binomials. The multifactorial analysis of constraints in two samples of corpus data (one for each group) yielded little evidence for the lexical unit hypothesis, as only one ordering constraint exhibited the predicted difference. In contrast, substantial evidence was acquired for the selection pressures hypothesis, as all ordering constraints yielded more pronounced effects in the sample of irreversibles. A more detailed exploration of this result revealed two important differences between the two samples: (i) ordering constraints are more strictly adhered to in irreversibles and (ii) ordering constraints affect a greater share of the data. In discussing these findings within the framework of language production research, I argued that the ordering constraints can be related to activation differences between the to-be-ordered constituents. Based on this I put forth the explanation that both of the observed differences can ultimately be explained by processing preferences. Processing ease may thus be considered to be a central factor underlying the emergence of irreversible binomials.

In terms of an outlook, it would be interesting to explore whether the processing argument can be generalized to other idiomatic constructions, i.e. addressing the question whether processing factors can be shown to explain the form of many (in principle all) lexicalized and/or idiomatic constructions.

References

Abraham, Richard D. 1950. "Fixed Order of Coordinates: A Study in Comparative Lexicography". MLA 34(4), 276-287.

Baayen, Rolf H. 2008. Analyzing linguistic data: A practical introduction to statistics using R. Cambridge: Cambridge University Press.

Benor, S. B., & Levy, R. 2006. "The chicken or the egg? A probabilistic analysis of English binomials". Language 82 (2), 233–278.

Bock, Kathryn J. 1982. "Towards a cognitive psychology of syntax: information processing contribution to sentence formulation". Psychological Review 89, 1–47.

Bock, Kathryn; Warren, Richard. 1985. "Conceptual accessibility and syntactic structure in sentence formulation". Cognition 21, 47–67.

Bock, Kathryn;Levelt, Willem J.M. 1994. "Language production: grammatical encoding". In Morton Ann Gernsbacher (ed.), Handbook of Psycholinguistics. San Diego, CA: Academic, 945-984.

Bolinger, Dwight. 1962. "Binomials and Pitch accent". Lingua 11, 34–44.

Cooper, W. E.; Ross, J. 1975. "Word order". In R. Grossman, R. E.; L. J. San & T. J. Vance (eds.). Papers from the Parasession on Functionalism April, 17, 1975, 63–111.

Crawley, Michael J. 2005. Statistics. An introduction using R. Chichester: John Wiley.

Dell, Gary S. 1986. "A spreading-activation theory of retrieval in sentence production". Psychological Review 93(3), 283–321.

Dell, Burger, Gary S., Lisa K.;William R., Svec. 1997. "Language production and serial order: A functional analysis and a model". Psychological Review 104(1), 123–147.

Dell, Gary S.; O'Seaghdha, Padraig G. 1994. "Inhibition in interactive activation models of linguistic selection and sequencing". In Dale Dagenbach; Thomas H. Carr (eds.). Inhibitory processes in attention, memory and language. San Diego/New York/Boston/London/Sydney/Tokyo/Toronto: Academic Press, 409–453.

Fenk-Oczlon, Gertraud. 1989. "Word frequency and word order in freezes". Linguistics 27, 517–556.

Kuiper, K., van Egmond, M.-E., Kempen, G.; Sprenger, S. 2007. "Slipping on super-lemmas: Multi-word lexical items in speech production". The Mental Lexicon 2(3), 313–357.

Lambrecht, Knud. 1984. "Formulaicity, frame semantics, and pragmatics in German binomial expressions". Language 60(4), 753–796.

Levy, Roger. in progress. Probabilistic models in the study of language. Cambridge, MA: MIT Press. (January 27, 2012 - http://idiom.ucsd.edu/~rlevy/textbook/text.html)

Lohmann, Arne. 2011. Constituent order in coordinate constructions – a processing perspective. PhD thesis. University of Hamburg.

McDonald, Janet L.; Bock, Kathryn J. & Kelly, Michael H. 1993. "Word and World Order: Semantic, Phonological, and Metrical Determinants of Serial Position". Cognitive Psychology 25. 188–230.

Malkiel, Yakov. 1959. "Studies in irreversible binomials". Lingua 8, 113–160.

Masini, Francesca. 2006. "Binomial constructions: Inheritance, specification and subregularities". Lingue e Linguaggio 5(2), 207-227.

Mollin, Sandra. 2012. “Revisiting binomial order in English: Ordering constraints and reversibility”. English Language and Linguistics 16(1), 81-103.

Müller, Gereon. 1997. “Beschränkungen für Binomialbildungen im Deutschen“. Zeitschrift für Sprachwissenschaft 16 (1), 25–51.

Norrick, Neal R. 1988. “Binomial meaning in texts”. Journal of English Linguistics 21 71–87.

Olsen, Susan. 2002. "Coordination at Different Levels of Grammar." In Marion Gymnich, Ansgar Nünning & Vera Nünning (eds.), Literature and Linguistics:Approaches, Models, and Applications: Studies in Honor of Jon Erickson, 169–188. Trier: Wissenschaftlicher Verlag Trier.

Peterson, Gordon E.;Lehiste, Ilse. 1960. “Duration of syllable nuclei in English”. Journal of the Acoustical Society of America 32(6), 693–703.

Pinker, Stephen;Birdsong, David. 1979. “Speakers' sensitivity to rules of frozen word order”. Journal of Verbal Learning and Verbal Behavior 18 (18), 497–508.

Ross, John R. 1982. “The sound of meaning”. In The Linguistic Society of Korea (ed.), Linguistics in the Morning Calm.Seoul: Hanshing Publishing Company, 275–290.

Sambur, Marnie. 1999. Factors that influence word ordering of conjunctive phrases containing a male and a female name: unpublished paper, available at http://bespin.stwing.upenn.edu/~upsych/Perspectives/1999/sambur.htm.

Steinlen, Anja K. 2002. A cross-linguistic comparison of the effects of consonantal contexts on vowels produced by native and non-native speakers. PhD thesis. Århus: English department of the University of Århus.

Wiechmann, Daniel; Lohmann, Arne (forthcoming). “Domain minimization and beyond: Modeling PP ordering”. Language Variation and Change.

Wright, S. K., Hay, J.; Bent, T. 2005. “Ladies first? Phonology, frequency, and the naming conspiracy”. Linguistics 43 (3), 531–561.

The Vienna English Language Test (VELT)

Susanne Sweeney-Novak, Vienna *

1. Introduction

Since shortly after the inception of a new curriculum at the Department of English at the University of Vienna in the autumn of 2002, a published standardized test had been used to assess first semester students at the beginning of their language competence course to establish their proficiency level in accordance with the *Common European Framework of Reference* (CEFR, Council of Europe 2001). Monitoring the results over time, it became clear that about 20 per cent of students did not meet the required B2 level, which is the level students leaving secondary education at age 18 are expected to have reached. It was felt at the time that students who did not reach this 'Matura' (= Austrian matriculation exam) level, i.e. B2 in the CEFR, should first improve their language competence through self -study before being offered a place in English language competence classes at the department.

Over the years, the level groups A1 (beginners) to C2 (very advanced) of the CEFR had shown consistency and stability in their size in the test data collected in this fashion.[1] However, issues of test security and related concerns led to the development of the Vienna English Language Test (VELT), which was first implemented in October 2011. This article will focus on quality assurance in the development, analysis and evaluation of VELT. The rules of constructing and selecting items, as well as statistical analysis and test theoretical methods of evaluation will be described.

* The author's e-mail for correspondence: susanne.sweeney-novak@univie.ac.at

1 Our findings of the distribution of proficiency in the student population are confirmed by extensive research at the Fachhochschule Wiener Neustadt (see Platzer 2010).

2. Describing VELT

It took about one-and-a-half years to develop VELT, although preliminary work had been done over the preceding years. It was agreed that VELT should in format and length be modeled on the standardized test previously used (in the following referred to also as 'the old test'), which had been a stable and robust measuring instrument using multiple choice items. The most important reason for this decision was practicability. With up to 800 students to be tested at the beginning of each semester, a test must be administered speedily and results produced quickly. Using answer sheets which can be machine read, it is now possible to have the results for a large number of students within a matter of hours.

Unlike the old test, which was a placement test and therefore targeted all the proficiency levels from A1 to C2 in the CEFR with the purpose of putting learners into more or less homogenous groups, the purpose of VELT is to establish whether a test taker is proficient in grammar and vocabulary at the level of B2 and above. In contrast to the old test, therefore, VELT only marginally includes items at B1 level and below.

VELT is administered to students wishing to embark on a course at the Department of English at the University of Vienna. It is a moderate-stakes test. This means that failing the test does not exclude a student from attending lecture courses at the department and is therefore not an entrance requirement for the undergraduate programme as such, but monitors access to the language competence courses. There is no limit to the times the test can be taken.

Two equivalent versions of VELT have been developed. They consist of 60 items each and time given for completion is 30 minutes. The format is four-options multiple choice throughout with one correct answer only. The decision to adopt a multiple choice (MC) format only rather than develop a test using a variety of formats was governed by the assumption that this is a format known to all test takers, thus avoiding the problem of test method influencing test performance. Furthermore, according to Purpura (2004), despite the criticism they receive, MC items are well suited for testing discrete features of grammatical knowledge. This claim would, no doubt, also apply for knowledge of vocabulary. In addition, items can be scored objectively (machine read), thus avoiding any subjective interpretation of student answers.

VELT consists of individual sentences and five short passages of between 70 to 90 words with between 7 to 9 gaps each. The passages include a range of text types. These texts provide opportunity to test grammar and vocabulary

beyond the sentence level, e.g. cohesion and coherence, logical connectors, and references.

3. Criteria for the quality of VELT – Validity

In order for a test to receive credence, its developers must ensure that it is valid and reliable. A test is valid when it serves a specified purpose; in the case of VELT, for example, to determine whether a student's lexico-grammatical proficiency is of the required level. For a test to be valid, test designers have to follow a theoretical framework and have to prove that the test only tests what it claims to be testing. The various aspects of validity as regards VELT will be discussed below.

3.1. Construct validity

For a test to have construct validity, it should be based on a theoretical framework. For VELT, the theoretical framework for testing grammar was adopted from Purpura (2004: 78), whose definition of grammar is represented in Table 1.

Grammatical form	**Grammatical meaning**	
Phonological/graphological forms (*not included*) Lexical forms co-occurrence word formation Morpho-syntactic forms tense and aspect word order (*not included*) mood Cohesive forms logical connectors cohesive devices Information management forms topic/comment (*not included*) Interactional forms hedging devices (*not incuded.*) backchannel devices (*not incl.*)	Phonolgical/graphological meanings Lexical meanings denotations connotations Morpho-syntactic meanings past time interrogation negation Cohesive meanings contrast, conclusion reference Information management meanings emphasis/foregrounding Interactional meanings (dis)agreement (un)certainty acknowledgement	Subsentential or sentential levels ↕ Suprasentential or discourse level
Low to High Context		

Table 1: Purpura (2004): Theoretical definition of grammar

In the table, those specific areas which are not included in VELT are indicated. In addition to grammatical form and grammatical meaning, Purpura includes a third column “Pragmatic meaning (implied)” which is not listed here because VELT does not include items which test this specific aspect of language.

The distinction between grammar and lexis is not straightforward, as was apparent when analysing and labelling individual items in VELT. Lexical forms and lexical meanings are listed in Purpura’s theoretical framework of grammar, which indicates that he sees grammar and vocabulary not as separate traits or single dimensions, but as being interrelated and the construct ‘grammar’ as being multidimensional.

Read (2000: 98) also mentions the problem researchers have in determining whether vocabulary and grammar are two distinct constructs, but he concludes that one of the problems in defining an item as testing vocabulary or grammar stems from the fact that the test method, the task, might influence the outcome and wrong inferences might be drawn from the results. VELT claims to test the lexico-grammatical proficiency of its test takers by including items in the areas listed in Purpura's model.

In VELT, vocabulary and grammar are tested context-dependently (see Read 2000: 9ff. on the dichotomy of context-dependent and context-independent). Every word or structure is part of a sentence which governs the meaning under scrutiny, or part of a text to test meaning beyond the sentence level. There is no separate grammar or vocabulary section.

The reading texts are not assessed by way of comprehension questions, but include selected cloze items which require the test taker to supply a missing word or phrase or grammar point chosen from four options. As has been mentioned before, the purpose of the reading texts is to test beyond the sentence level, and to test text specific features such as past tense or participle clauses or logical connectors.

3.2. Content validity

A test can claim to have content validity if the tasks are designed in such a way as to produce evidence for what they set out to measure. Basically, evidence is collected to answer the following question: Which inferences can be drawn from the test scores to the test takers? VELT aims to ensure that test takers are proficient at recognizing the correct use of core concepts of English morphology and syntax and can provide evidence of mastering vocabulary up to the 3000 word level (cf. Nation 2001), but contrary to many vocabulary tests, which ask for definitions or elimination of non-words, VELT focuses on the meanings of words, semantic fields and collocations.

VELT is not able at present to make any statements about a test taker's productive abilities, for instance how well a student can write texts. However, research has linked vocabulary knowledge to reading ability. Alderson (2000: 99) states that "[t]ests of vocabulary are highly predictive of performance on tests of reading comprehension." It could, therefore, be assumed that those students who do well on the VELT might also be good readers of academic texts.

3.3. Predictive validity

We have no information to date which would link VELT scores to overall academic achievement. However, on two occasions, scores from the old test used initially were compared with results/marks at the Common Final Test (CFT), a standardised test developed at the Department of English at the University of Vienna. Students take the CFT at the end of their second semester. The CFT is a comprehensive test of reading and writing but has no separate grammar or vocabulary section. There was a high correlation between scores received on the standardised test and the CFT. No comparison has been undertaken between scores and semester grades as the latter was felt to be subject to too many variables such as homework, oral presentations, etc.

4. Developing VELT

4.1. Phase 1: Collecting data

'Language in use' items from published test papers were given to students at the beginning of their first semester so that no learning could have taken place between leaving secondary school and entering university. The purpose of using papers from published tests was to gain an understanding for the level of proficiency students are at and which types of item are typical of a specific level of proficiency. At the same time, independently constructed items were piloted to see whether these correlated with standardized items at specific levels of proficiency.

Two versions of a trial test using parallel items to the published standardised test used previously were developed and items from the trial run which proved appropriate in terms of difficulty and discrimination were incorporated. Lexical items were included which were taken from the

Academic Vocabulary List (cf. Nation 2001),[2] which had been used for assessment and teaching in many previous semesters together with the 2000, 3000 and 5000 word lists to assess students' vocabulary levels for remedial purposes.

All short reading texts are authentic texts taken from different sources. Sometimes it was necessary to make minor adaptations for the text to fit the proficiency level it aimed at, to ensure that the content was not biased or world knowledge required to understand the content, and to be in keeping with the required length of these short passages.

In selecting items, a preference was given to British English though care was taken not to discriminate against speakers of American English. A considerable part of the discrete sentences was taken from the British National Corpus or other corpora or from dictionaries whose examples are based on a corpus. However, in some rare cases at the lower level of proficiency, it was easier to construct an item independently to cover the point under consideration.

4.2. Phase 2: Including and excluding items

Items which looked appropriate from the point of view of item difficulty and discrimination (Tables 2 and 3) were piloted. Distractor analyses (Table 4), i.e. an analysis of the incorrect choices incorporated into the MC test, were conducted and adjustments of weak distractors made. After analysis, two versions of the test were calibrated.

This is an example (from trial version 2) of a first rough analysis of the efficacy of each item: difficulty level (= facility value) and discrimination index (how well an item distinguishes between more or less able candidates). For this analysis SPSS Version 17.0 was used.

Cronbach's Alpha	Number of Items
.800	62

Table 2: Reliability statistics

Table 2 indicates that the reliability of the trial test is .80, which, for a first draft is quite acceptable. The higher, of course, this figure would be, the better. The number of items under review is given.

[2] The Academic Word List was developed by Averil Coxhead, University of Wellington, New Zealand. It contains semantic fields specifically apparent in academic texts.

Table 3 tells us the facility value of each item (in SPSS the facility value is labelled 'mean'): a facility value of .99 says that 99% of the candidates answered the item correctly. Easy items of .8 or .9 are at the beginning of the test to 'relax' students and also to target low proficiency candidates.

	Mean	Standard Deviation	Corrected item-total correlation	Cronbach's Alpha if item deleted	N
item1	.9947	.07274	.065	.800	189
item2	.8942	.30842	.072	.800	189
item3	.7566	.43027	.018	.803	189
item4	.4392	.49760	.077	.802	189
item5	.9312	.25376	.100	.800	189
item6	.3915	.48939	.146	.800	189
item7	.5185	.50098	.262	.796	189
item8	.8519	.35619	.148	.799	189

Table 3: Item statistics

Table 3 also indicates how statistical information (most importantly the Reliability Index Cronbach Alpha) on a test would change if an item were to be discarded. It also informs us of the discrimination value of an item, listed in the column 'Corrected item-total correlation'. The discrimination index states how successfully an item discriminates between proficient and less proficient test takers (with items of high facility value, discrimination is less likely). Item 7, for example, seems to be working well: it has a desired facility value of .51, it discriminates quite well, and if we were to discard it, the reliability index would be lowered. On the other hand, if we look at items 3 and 4, we see that item 3 with a facility value of .76 is fairly easy; we can also see that it does not discriminate between high and low level proficiency candidates as the discrimination index of .018 is below the required .3+ , and that the reliability index would rise slightly if this item were to be removed from the test. As a last piece of information in the item analysis process, we would inspect how well or badly the distractors for each item worked. This means looking at how attractive, over-used or under-used the individual distractors were.

Despite its ideal facility value (around .50), and fairly acceptable results in the distractor analysis (except for distractor C), item 4 does not discriminate (Corrected item-total correlation = .077) and is in need of change.[3]

		Frequency Distribution	Per Cent	Valid Per cent
Valid	A	49	25.9	25.9
	B=key	85	45.0	45.0
	C	4	2.1	2.1
	D	51	27.0	27.0

Table 4: Distractor Analysis: Frequency Distribution

In addition to omitting items which were found statistically wanting, some items were discarded because of the feedback received from students and colleagues. These included items which showed bias, for example discriminating against speakers of American English or items which did not take into consideration language change.

4.3. Phase 3: Trialling

Two versions, referred to as Version 1 and Version 2 below, of the future test were trialled and correlation studies with the old entrance test were conducted. 189 first-semester students, who had previously taken the old test, took both versions of the new test.

In many cases there was an ideal match between students' results received on the old test and the results on both trial versions, in some cases there were discrepancies, and in some cases one version reported much better results than the other. The fact that there were discrepancies in some cases between trial versions 1 and 2 could have been due to well-known factors which influence reliability, e.g. students arriving late or leaving early, thus not doing all items. In addition, having been told at the outset that the results of the trial would be of no consequence for them, some students might not have taken the trial too seriously. Before the trial sessions, students were told that these tests should give their teachers some ideas for remedial teaching.

[3] It is interesting to note though that both items 3 and 4 were grammar items testing the use of the present tense for future aspect and testing the use of the present perfect respectively. It could well be an interesting research project to see whether in a grammar and vocabulary test there is a difference in the efficacy of items which test knowledge of grammar and those which test knowledge of vocabulary.

5. Validating the trial versions

5.1. External validity: Concurrent validity using correlation studies

Having used the old test for entrance purposes for a number of years, and having found the results of the test to be consistent and sound in determining which students are at the required ability level, it was clear that this test should be used to establish the new test's concurrent validity by way of correlation studies.

The greatest problem for an accurate analysis of the trialled versions was the fact that the first-semester students represented only 50% of the overall population which had originally taken the old test before the start of the summer term 2011. These ranged from B2+ to C2 in the CEFR. This meant that we had little information on how students of lower proficiency would manage the new test.

There were three sets of data to work with. The 189 students in the trial had taken the old test before the start of the semester and had subsequently taken both versions of the new test. It was therefore possible to correlate the two new test versions with an external measurement instrument, namely the old test. First of all, the relationship between variables was checked by using scatterplots.

The scatterplots (see Appendix 1), show a positive relationship between three variables. This is indicated by the lines which run from the bottom left-hand corner to the top right-hand corner. Such a line is referred to as the 'fit line' or the 'line of regression'. The positive relationship indicates that as performance increases on one variable, so does performance on the other. For our purpose this would indicate that a student taking one test and showing a specific result might show a similar result taking one of the other tests under scrutiny. Furthermore, there is a stronger relationship at the lower range of the scores where there are greater clusters. However, there are outliers along all lines, but more so in the first scatterplot (Version 1 vs. old test) and at all the upper ranges of the scores. These indicate that there are test takers who do not score consistently high or low in both variables and could be an indication of the points made above about influences on reliability.

As a further step to see to which extent two sets of data agree with each other, the *Pearson product- moment correlation coefficient* (referred to as r or R) was calculated. As the scatterplots showed some outliers, those data with significant discrepancies were taken out of the data set: for example, more than 10 points difference between trial and the old test and cases with high numbers of items missing. In these instances it was not clear whether students

had arrived late to do the trial test, whether they had not taken the trialling process seriously enough, or whether test security on the old test had been compromised. The remaining number of cases was 138.

The correlation indices (see Appendix 2) show that there is quite a strong agreement between the old test and the trial versions 1 and 2 (.794 and .761 respectively), as well as between trial version 1 and trial version 2 (.809). The fact that correlation indices between both versions and the old test are lower could probably be due to the already mentioned differences between the old test and the trial versions, i.e. a higher number of grammar items in the trial versions and slightly longer reading passages.

To see whether and to what extent trial version 1 and trial version 2 measure the same construct, an *r* of .809 is acceptable although an *r* in the high .80s or .90s, according to Hatch & Lazaraton (1991: 440f), would be desirable. *r* squared = .65 indicates that there is a 65 per cent overlap between the two versions. The unique variance of 35 per cent means that 17.5 per cent are unique to trial version 1 and 17.5 per cent to trial version 2. This overlap indicates that two thirds of both versions measure the same construct, but to determine the nature of the unique variance is fairly difficult.

5.2. Validating the trial versions using Item Response Theory (IRT)

So far in the article, statistical information quoted has been part of what is termed 'Classical Test Theory' (CTT). In the following, aspects of modern test theory will be included and data will be presented based on an analysis using Item Response Theory.[4] IRT, it is claimed, is superior to CTT as it is a probabilistic measurement theory. IRT, in comparison to classical test analysis, provides test developers with additional information about test takers as well as items. It is a powerful statistical tool which is used to make informed claims about a test's overall quality, about item and person characteristics and about their relationship. IRT models are based on formalized expectations about person and item behaviour which is not directly observable; hence, IRT models are also referred to as 'latent trait' models. The theory states that performance on an item reflects a candidate's level of ability in relation to item characteristics (Bachman 2004:141).

In IRT, a number of models can be used. In the trial and subsequently in the VELT analysis, the one-parameter Rasch model was used to determine levels of proficiency and to determine which items fit the Rasch model and

4 The following books provide excellent introductions to IRT: Henning (1987), McNamara (1996), Bond & Fox (2007).

match person ability. The software used for the Rasch analysis was WINSTEPS 3.70.0.

Figure 1 on the following page depicts the results of Trial Version 2 and should serve as an example to illustrate how Rasch analysis can be a vital part of test development and interpretation. On a descending scale from +4.6 to -4.4 logit units the relationsh ip between person ability and item characteristic is drawn. At the very top of the person column there is a dot representing one test taker and at logit 4 there is the same information: one test taker on the same logit as item 61. We know about these two test takers that the first has a 62 per cent chance of answering the item correctly and that the second one has a 50 per cent chance of doing so. The third test taker at 3.6 logits has a 38 per cent chance of answering the item correctly. Test takers at 3 logits have a 27 per cent chance.

The average item difficulty is, by default, set at 0, which in the map below means that item 54 is of average difficulty and items 12, 29, 3, and 50 are just below average. In the trial versions, the average test taker's ability in relation to item difficulty is at +1 logit. From Figure 1 it can immediately be observed that about 20 items are below -1 logits and below the weakest test taker, meaning that these items do not fit the population. It can be assumed that these twenty items target lower ability levels up to B2-. Furthermore, the majority of the test takers are above the item average of 0 logits. This is a reflection of the fact that first-semester students in the summer term 2011 represented only 50% of the overall proficiency of candidates who sat the old test. If Figure 1 had represented the whole range of proficiency, we might have seen a shift of items to 0 logit, and probably also a greater number of test takers around 0 logit and below.

Figure 1: Person/Item Mapping Trial Version 2. Each "#" is 2. Each "." is 1.

```
INPUT: 189 PERSON  62 ITEM  MEASURED: 189 PERSON  62 ITEM  2 CATS WINSTEPS 3.70.0.2
--------------------------------------------------------------------------------
           PERSON - MAP - ITEM
              <more>|<rare>
   5                +
                    |
                    |
                .   |
                    |
                    |
   4             .  +  ITEM61
                    |
                .   |
                    |
               .#   |
                    |T ITEM60
   3                +
                . T|
                .   |  ITEM55
              .##   |  ITEM21
              ###   |
               .#   |  ITEM53
   2           .# S+  ITEM27
             ####   |  ITEM30  ITEM57
            ######  |  ITEM36  ITEM6
             #####  |S ITEM24  ITEM62  ITEM63
      #############  |  ITEM4
            ###### M|  ITEM25
   1   .##########  +  ITEM19  ITEM7
       .##########  |  ITEM33  ITEM37
            .#####  |  ITEM14  ITEM52  ITEM64
         .########  |  ITEM20  ITEM31  ITEM39  ITEM40
            .##### S|  ITEM32  ITEM56
              .###  |
   0             #  +M ITEM54
                 .  |  ITEM12  ITEM29  ITEM3   ITEM50
                 #  |  ITEM17  ITEM45
                ## T|  ITEM10  ITEM22  ITEM26  ITEM41  ITEM49
                    |  ITEM15
                 .  |  ITEM18  ITEM34  ITEM51  ITEM8
  -1                +  ITEM35  ITEM48
                    |  ITEM16  ITEM2
                    |  ITEM23  ITEM38  ITEM47  ITEM58
                    |S ITEM59
                    |  ITEM28  ITEM43  ITEM5
                    |  ITEM46
  -2                +
                    |
                    |
                    |  ITEM11  ITEM13
                    |
                    |
  -3                +
                    |T
                    |
                    |
                    |  ITEM9
                    |
  -4                +
                    |
                    |  ITEM1
                    |
                    |
                    |
  -5                +
              <less>|<frequ>
```

5.3. Standard setting

One of the most challenging steps in the development of VELT was addressing the question of the cut scores between CEFR levels, and, most specifically, setting the cut score between pass and fail. Various paths were taken to tackle the problem. First of all, the difficulty (facility value) of an item gave some indication whether an item was easy or difficult. As we had the results of the old test, we were able to match test takers' CEFR levels based on the old test with their scores on the trial versions. We also used the IRT person/item map (see section 5.2.) together with the facility values to determine the cut points between levels.

As a second step and in an attempt to follow the guidelines in the manual *Relating Language Examinations to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR)* (Council of Europe 2009), colleagues who also teach at grammar schools with many years of experience and frequent 'Matura' classes were asked to scrutinize the items and decide whether a student at B2 = 'Matura' level would be able to answer an item correctly, whether they would regard an item as below 'Matura' level, or whether only advanced students would be able to answer an item correctly. This is, of course, not the classic standard setting procedure as described in the literature and specifically in the manual, but given the fact that standard setting is complicated and needs training, this was the best alternative. Besides, there were only very few items on which colleagues did not agree. However, some discrepancies were found between colleagues' judgments and item difficulty, i.e. the facility value.

As a third step, the *English Vocabulary Profile* was consulted. This instrument, which has been published on the web fairly recently, lists lexical items together with their associated CEFR levels. There were some discrepancies between our data (facility values and teacher assessment) and the *English Vocabulary Profile* results. For example, one of our early items tested the knowledge of '*had better'* with a high facility value for '*You'd better hurry*'. This coincided with the profile's indication of this vocabulary item/structure being at B1. However, a second item testing the same phrase but in a different context showed that this structure is not even mastered at C2. The item again was testing the *'had better'* phrase, asking for the tag to be supplied: '*You'd better hurry, hadn't you*?'. The option most students chose was '*shouldn't*', showing that they can form a tag question but do not

know what the ‘ …’*d* ’ stands for. With a facility value of .07, this item had to be omitted from the test.[5]

As a fourth step, we drew on the results from the previous 16 semesters about the distribution of proficiency levels of beginning students. This gave us a good idea as to which percentage of test takers would be below B2 and which in the B2, C1 and C2 range. Whether these four approaches to standard setting are indeed reliable will be evident when VELT is administered and analysed in the future.

6. Validating the final product VELT: classical test analysis and item response theory

When VELT was administered for the first time, 621 candidates took the test. They were split randomly into two groups: Group 1 (n= 333) took Version 1 and Group 2 (n= 288) took Version 2. The most important questions to be asked are: Are the two versions of the test equivalent? How reliable are the statistical findings? These questions will be addressed in the following section.

6.1. Comparability of the two groups

An independent-sample t-test was used to test the hypothesis that the differences in the mean between group 1 and group 2 are due to chance. For a t-test to be meaningful, certain criteria have to be met. These are adapted from Hatch & Lazaraton (1991):

i. There are only two levels (groups) of one independent variable to compare.

ii. Each test taker is assigned to only one group.

iii. The data are truly continuous.

iv. The mean and the standard deviation are the most appropriate measure to describe the data.

v. The distribution is normal and the variances are equivalent.

[5] From a language acquisition point this might be quite interesting. This structure appears early in course books, but is obviously never explored or just forgotten. A phrase like “*best be going*” or “*we’d better be going*”, which is frequently heard in spoken BE, is apparently not part of students’ active vocabulary even at an advanced stage.

The data meet these criteria with a normal distribution: 2 SD fit on either side of the mean in groups 1 and 2, and the test overall. An independent-sample t-test was conducted to compare the two groups of the test.

Group	N	Mean	Standard Deviation	Standard Error of the Mean
1	333	36.04	9.832	.539
2	288	37.50	9.386	.553

Table 5: Group Statistics

	Levene's Test for Equality of Variance		T-Test for Equality of Means						
								95% Confidence Interval of the Difference	
	F	Sign.	t	df	Sign. (2-tailed)	Mean Difference	Std. Error of Difference	Lower	Upper
Equal variances assumed	.869	.352	-1.886	619	.060	-1.461	.775	-2.983	.060
Equal variances not assumed			-1.893	612.975	.059	-1.461	.772	-2.978	.055

Table 6: Independent Sample T-Test

The significance value of the Levene-Test (p = .352) being larger than .05, equal variance can be assumed; **t = -1.89, p = .06 (two-tailed)**. The results of the t-test show that there was no significant difference in scores for group 1 (mean = 36.04, standard deviation = 9.83) and group 2 (mean = 37.5, standard deviation = 9.39). The significance value (two-tailed) of .06 is not significant at the critical value of .05, indicating that there is not sufficient evidence to reject the null hypothesis (differences in the means are due to chance).To gauge the size of association or difference in the groups, should it nevertheless exist in the population, the **effect size** was also calculated. Basically, this is a simple way of quantifying the size of the difference between two groups. One formula[6] to calculate effect size is Cohen's *d*,

[6] Formula: Effect size = Mean of one group minus mean of other group divided by standard deviation.

which gives us an effect size of .15, i.e. rather small by most estimates. The other is eta squared[7] = .0057; expressed as a percentage, only 0.57 per cent of the variance can be explained by the group in which test takers find themselves.

6.2. Reliability

SPSS provides a number of statistical procedures whose results can help to make inferences about the quality of a test. Appendix 3 displays the number of test takers when VELT was first administered (N valid and N missing). In addition, there are the figures of central tendency: mean, median and mode. In this test, these figures are close together, except for Version 1, where we have two modes, 32 and 36. Skewness and kurtosis figures indicate the shape of the distribution: within the range of +2 to -2 the score distribution can be regarded as normal (Bachmann 2004: 74). Skewness figures are low. Dividing skewness by its error shows a figure of .156 for VELT version 1 and .694 for VELT version 2. As this is within the range of -2/+2 this is acceptable at the .05 significance level. Whether there is a problem with kurtosis can be seen using the same procedure: VELT version 1 = .162 and VELT version 2 = .186. Skewness and kurtosis figures should be inspected to ensure normal distribution.

Cronbach Alpha:	Number of items:	Standard Error of Measurement:
Version 1 .894 Version 2 .888	60 each	3.14[8]

Table 7: Cronbach Alpha and Standard Error of Measurement

Table 7 provides the Cronbach Alpha reliability index, which in this case is .894 on a scale from 0 to 1, and can be regarded as a satisfactory figure. McNamara (2000: 58, 62) recommends a reliability index of .9 or better depending on what is at stake. Table 7 also shows an acceptable standard error of measurement of 3.14 (QPT = 4). This means that we can be 68 per cent confident that the true score of a test taker is within +/- 3 scores of the

[7] Pallant (2007: 235f): eta squared gives additional information regarding the strength of association between the independent variable (group) and dependent variable (result). Formula for eta squared and the website for Cohen's *d* are found in Pallant (2007).

[8] Formula taken from Hughes (1989: 159).

raw score. For example, a candidate scoring 45 points on the test would fall somewhere between 42 and 48 points.

6.3. Item Response Theory (IRT)

The one-parameter Rasch analysis (as discussed in section 5.2) was also used in the analysis of VELT. Bachman (2004: 142) stresses how important it is when using Rasch analysis that "an IRT model fits the data" (see explanation below under 'Standardized Residuals'). Ensuring fit is, of course, essential to implementing a test. Appendix 4 is a copy of the results gained from WINSTEPS as regards fit statistics for VELT versions 1 and 2. Those figures which are important for the interpretation of the VELT data are discussed below:

Standardized Residuals (the difference of what is expected by the Rasch model and what is observed) indicate that in both versions the **mean** is zero and the **standard deviation** is close to 1 (1.01/1.00). This is exactly what the figures should be, because they indicate that the data conforms to the basic Rasch model specifications.[9]

Person Reliability in both versions is .89. This figure indicates the extent to which the person ordering would be replicated if this sample of persons were to be given another set of items measuring the same construct. As with all figures of reliability, the more the figure approaches 1 the better. The figure of .89 (which corresponds to SPSS's Cronbach Alpha) is satisfactory, meaning that we can trust the results rendered by the test.[10]

Item Reliability in version 1 is .99 and in version 2 it is .98. On a scale from 0 to 1, this is a very good result. In comparison with person reliability, it says that we have better information about the items than about the candidates. This could, for example, mean that there are test takers who were not challenged enough by the test.

The higher number of test takers (n = 621) as compared to the trial (n = 189), affected the mean scores of both VELT versions. It is the closeness of the figures in the trial version 1 and 2 which is striking in contrast to the figures in VELT as indicated in Table 8.

9 The trial versions showed very similar results: mean = .00 in both versions, SD (version 1) = 1.01 SD (version 2) = 1.00; Item Reliability was .98 and .97 respectively.

10 Person Reliability increased considerably from the trial where we had .81 for both versions. This increase is most probably due to the actual test population's spread across all ability levels in contrast to the trial population which, as has been mentioned above, represented only the top 50% of allability levels.

	Trial Version 1	SD	Trial version 2	SD	VELT Version 1	SD	VELT Version 2	SD
Person mean score	42.7	7.0	43.1	6.9	36	9.8	37.5	9.4
Item mean score	126.0	45.2	127.4	45.1	200	71.1	180	63.2

Table 8: Comparison of mean scores: trial versions and VELT versions

There are two possible answers to the question why we have this difference in the mean scores between trial and VELT: one is the fact that in the trial, as said before, only the top 50% of the proficiency range of prospective first-semester students were trial candidates; the second indicates that in VELT fewer items (approximately only ten items) targeted low proficiency candidates.

Figure 2 on the following page depicts the results of the person/item map of VELT version1. In contrast to Figure 1 (section 5.2), which represented the person/item map of one trial version, there is a clearer picture of the overall test-taking population. The average person ability is at 0.6 logits (trial 1 logit) and there is a considerable number of items at 0 logits.

Maybe it could be said that the test takers at the very top of the scale are not challenged enough and a couple of items more up the scale would have been desirable. At the other end of the scale, a couple of items are too easy. On the other hand, over 50 per cent of the items cluster around the +1 to -1 logits and 92 per cent range from +2 to -2 logits, indicating a sufficient spread along the ability range. Besides, the main focus of the analysis for departmental purposes is on the cut score between B1 and B2, which is around the 0 logits point, rather than attributing the whole student population to particular levels of the CEFR as one would do in a placement test.

	Range of person ability in logits	Range of item difficulty in logits
VELT version 1	4.60 to - 1.9	2.44 to -3.01
VELT version 2	4.10 to - 2.10	2.93 to -3.47

Table 9: Comparing range of person ability and item difficulty in the two VELT versions

On a final note, it seems from Table 9 that some test takers taking version 1 could have been more proficient than those taking version 2. As regards item difficulty, in version 2 only one item, item 54, is placed at 2.93 logits. As the next item is placed at 2.40 logits, a more equivalent picture arises.

Figure 2: IRT Person – Item Map. Each "#" is 3. Each "." is 1 to 2.

```
INPUT: 333 PERSON 60 ITEM  MEASURED:333 PERSON  60 ITEM  2 CATS WINSTEPS 3.70.0.2
--------------------------------------------------------------------------------

               PERSON - MAP - ITEM
                   <more>|<rare>
        5                +
                         |
                         |
                      .  |
                         |
                         |
                         |
                         |
        4             .  +
                         |
                         |
                      .  |
                         |
                      #  |
        3             .  +
                         |
                      # T|
                     .#  |T ITEM48
                     .#  |  ITEM47
                     .#  |
        2         .#####  +  ITEM29  ITEM56
                      .  |  ITEM49
                     .# S|  ITEM31  ITEM38
                   .######  |  ITEM50  ITEM52  ITEM53
                   .#####   |  ITEM36  ITEM58
                   .#####   |S ITEM24  ITEM32  ITEM54  ITEM60
        1             .##  +
                .########   |  ITEM25
                 .####### M|  ITEM20  ITEM59
                .########   |  ITEM21  ITEM42
                  #######   |  ITEM33  ITEM8
                #########   |  ITEM13  ITEM27  ITEM35  ITEM5   ITEM7
        0           .####  +M ITEM10  ITEM37  ITEM55  ITEM57
                 .#######   |  ITEM15
                   .#####   |  ITEM30
                   ###### S|  ITEM14  ITEM39  ITEM40  ITEM41  ITEM43
                     ####   |  ITEM16  ITEM19  ITEM45  ITEM51
                       ##   |  ITEM22  ITEM6
       -1               #   +  ITEM23  ITEM26  ITEM3   ITEM46
                      .#   |S ITEM18  ITEM28  ITEM9
                           |  ITEM34  ITEM44
                        # T|
                           |  ITEM1
                      .#   |
       -2                  +  ITEM17  ITEM2
                           |  ITEM11
                           |  ITEM12
                           |T
                           |
                           |
       -3                  +  ITEM4
                   <less>|<frequ>
```

7. Conclusion

The article has attempted to summarize the work that was done over a considerable length of time in order to develop an instrument to measure the proficiency level of beginning students of English at the Department of English, University of Vienna, and to ensure that they meet the required B2+ level of the Common European Framework of Reference, which is assumed to be the prerequisite for studying English in an academic context. The assumption was made that a test targeting lexico-grammatical knowledge would be a valid instrument for making inferences about the overall language ability of test takers for the purpose of studying in an English-speaking academic environment. For the department's purposes it was essential to develop a test which could indicate with confidence whether a student would meet the required proficiency level or whether more work must be done by some students to improve their language ability in English.

From the developer's perspective, the content of the test is appropriate for the target population on a suitable level of difficulty. This article has reported on the steps that were taken from the original conceptualization to piloting and trialling of prospective versions of the test, and finally to the implementation of VELT. The article has attempted to show why it would be justified to claim that VELT is a reliable and valid instrument based on sound test theoretical principles and thorough statistical analysis on the basis of which decision regarding students' further development in English could be made.

For the future, similar analyses should be undertaken to have further proof of the stability of this measuring instrument. Using IRT, perhaps an item bank could be established. A closer inspection of items testing grammar and items testing vocabulary and their influence on overall scores could also be a worthwhile research project.

Acknowledgements

Special thanks must be given to those colleagues who participated in piloting and trialling; to those who spotted errors in design, in accuracy, in usage; to those who commented critically on the usefulness and purpose of early items. I am especially grateful to friends and colleagues who participated in our attempt at standard setting, and to those colleagues who read drafts of the test and who helped me with statistical analyses, both at the Department of English, University of Vienna, and the Fachhochschule Wiener Neustadt. I acknowledge that parts of the article might be complicated and tedious to read for someone who is only marginally familiar with statistical data. For this

reason, my special thanks go to those colleagues at the department who read the article and provided me with an invaluable amount of feedback.

References

Alderson, J. Charles. 2000. *Assessing Reading*. Cambridge: CUP.

Alderson, J. Charles; Clapham, Caroline; Wall, Dianne. 1995. *Language Test Construction and Evaluation*. Cambridge: CUP.

Bachman, Lyle F. 2004. *Statistical Analysis for Language Assessment*. Cambridge: CUP.

Bachman, Lyle F.; Palmer, Adrian S. 1996. *Language Testing in Practice*. Oxford: OUP.

Bond, Trevor G.; Fox, Christine M. 2007. *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*. (2nd edition). New Jersey: Lawrence Erlbaum Ass.

Council of Europe. 2001. *Common European Framework of Reference for Languages: learning, teaching, assessment*. Cambridge: CUP.

Council of Europe. 2009. *Relating Language Examinations to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR): a manual*. Language Policy Division, Strasbourg. http://www.coe.int/t/dg4/linguistic/Source/ManualRevision-proofread-FINAL_en.pdf (1 July 2011).

Davies, Allen; Brown, Annie; Elder, Cathie; Hill, Kathryn; Lumley, Tom; McNamara, Tim. 1999. *Dictionary of language testing*. [*Studies in Language Testing 7*]. Cambridge: CUP.

English Vocabulary Profile. http://www.englishprofile.org/index.php?option=com_content&view=article&id=4&Itemid=5 (15 Sept. 2011).

Hatch, Evelyn, Lazaraton, Ann. 1991. *The Research Manual: Design and Statistics for Applied Linguistics*. Massachusetts: Heinle & Heinle.

Henning, Grant. 1987. *A Guide to Language Testing: development, evaluation, research*. Cambridge, Mass.: Newbury House.

Hughes, Arthur. 1989. *Testing for Language Teachers*. Cambridge: CUP.

Lienert, Gustav A.; Raatz, Ulrich.1998. *Testaufbau und Testanalyse*. (6th edition). Beltz, Weinheim, Basel.

McNamara, Tim. 1996. *Measuring Second Language Performance*. London and New York: Longman.

McNamara, Tim. 2000. *Language Testing*. Oxford: OUP.

Nation, I.S.P. 1990. *Teaching and learning vocabulary*. New York: Newbury House.

Nation, I.S.P. 2001. *Learning Vocabulary in Another Language*. Cambridge: CUP.

Nunan, David. 1992. *Research Methods in Language Learning*. Cambridge: CUP.

Platzer, Hans. 2010. "Educational standards in EFL and their attainability: An Austrian case study". *e-FLT (Electronic Journal of Foreign Language Teaching)* 7, 49-65. http://e-flt.nus.edu.sg/v7n12010/platzer.pdf (10 Nov. 2011).

Purpura, James E. 2004. *Assessing Grammar*. Cambridge: CUP.

Read, John. 2000. *Assessing Vocabulary*. Cambridge: CUP.

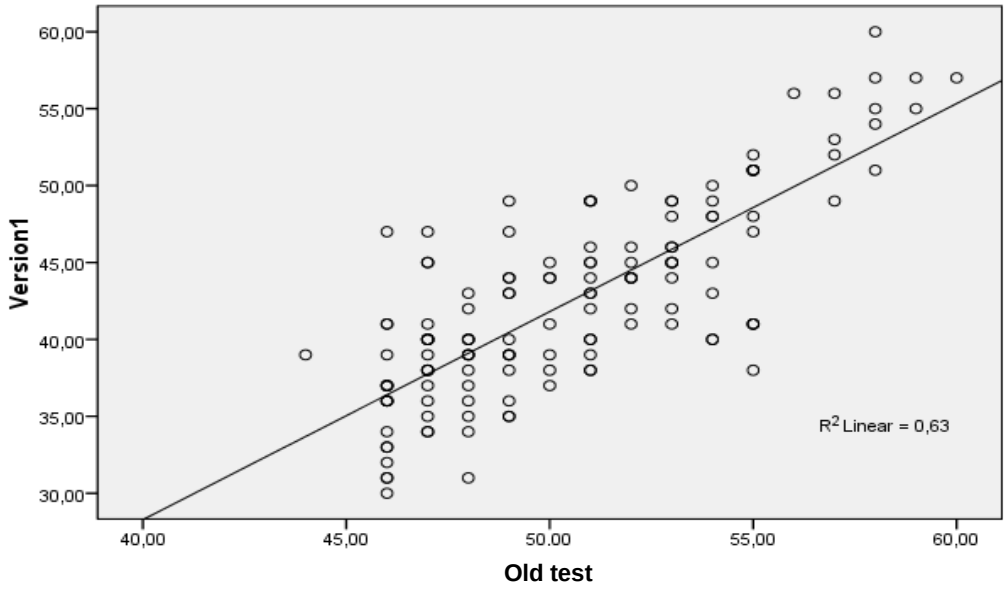
SPSS 17.0. Computer programme. http://www-01.ibm.com/software/analytics/spss/ (17 December 2012).

WINSTEPS 3.70.0. Computer programme. Director: Linacre, Mike. www.winsteps.com (1 July 2010).

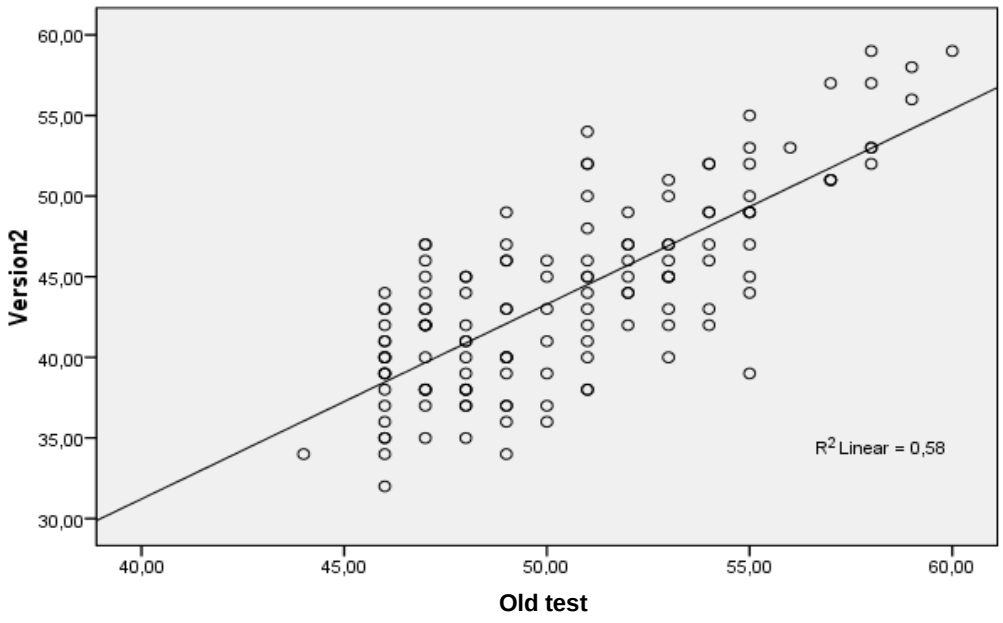
Appendices

Appendix 1: Three scatterplots comparing the results on the old test with the results on the two trial versions of the new test.

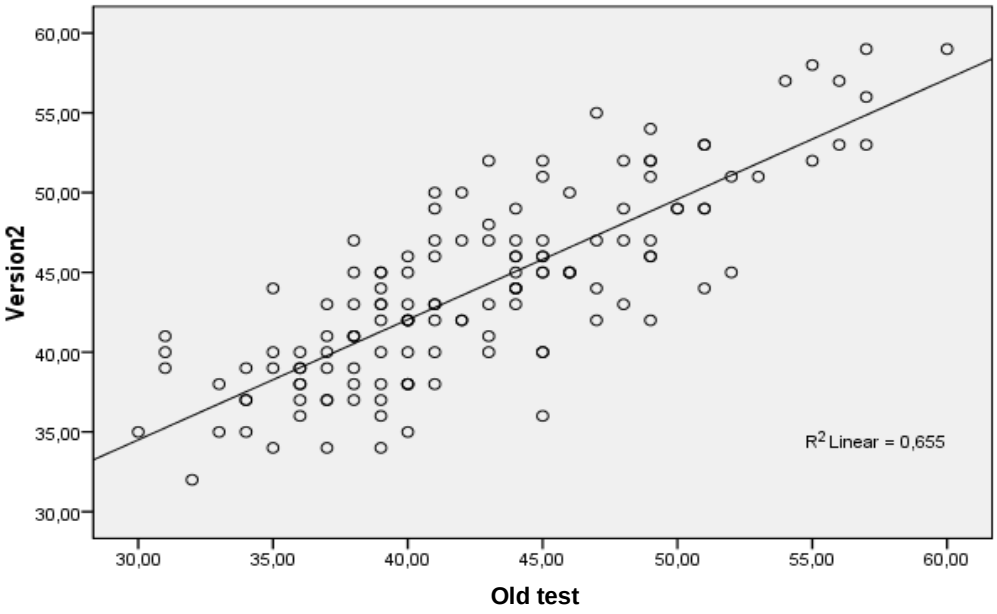
Plot 1: Trial Version 1 vs. old test. R squared = .63



Plot 2: Trial Version 2 vs. old test. R squared = .58



Plot 3: Trial Version 1 vs Trial Version 2. R squared = .66



Appendix 2: Correlations

Correlation of Version 1 and Old Test

		Version1	Old Test
Version1	Pearson Correlation	1	.794**
	Significance (2-tailed)		.000
	N	138	138
Old T.	Pearson Correlation	.794**	1
	Significance (2-tailed)	.000	
	N	138	138

**. Correlation is significant at 0.01 (2-tailed)

Correlation of Version 2 and Old Test

		Old test	Version2
Old T.	Pearson Correlation	1	.761**
	Significance (2-tailed)		.000
	N	138	138
Version2	Pearson Correlation	.761**	1
	Significance (2-tailed)	.000	
	N	138	138

Correlation of Version 1 and Version 2

		Version2	Version1
Version2	Pearson Correlation	1	.809**
	Significance (2-tailed)		.000
	N	138	138
Version1	Pearson Correlation	.809**	1
	Significance (2-tailed)	.000	
	N	138	138

Appendix 3: Descriptive Statistics

	VELT Version 1	**VELT Version 2**	**VELT total**
N Valid	333	288	621
N Missing	0	0	0
Mean	36.0420	37.5035	36.72
Std.Error of Mean	.53880	.55310	.387
Median	36.0000	37.0000	37.00
Mode	**32 / 36** [a]	37	37
Std.Deviation	9.83224	9.38640	9.648
Variance	96.673	88.105	93.083
Skewness	-.021	.100	.020
Std.Error of Skewness	.134	.144	.098
Kurtosis	-.432	-.532	-.451
Std. Error of Kurtosis	.266	.286	.196
Range	48.00	48.00	49
Minimum	11.00	10.00	10
Maximum	59.00	58.00	59

a. There is more than one mode: 32 points reached by 17 candidates; 36 points reached by 15 candidates

Appendix 4: Fit statistics

Calculating Fit Statistics using the 1-paramenter Rasch model

>====================================<

Standardized Residuals N(0,1) **Mean: .00 S.D.: 1.01**

Time for estimation: 0:0:0.296

VELT Version 1

PERSON	333 INPUT	333 MEASURED		INFIT		OUTFIT	
	SCORE	COUNT	MEASURE	ERROR	IMNSQ	ZSTD	OMNSQ	ZSTD
MEAN	36.0	60.0	.62	.33	1.00	.0	1.02	.1
S.D.	9.8	.0	1.03	.07	.14	.9	.38	1.0
REAL RMSE	.34	TRUE SD	.98	SEPARATION	2.86		**PERSON RELIABILITY**	**.89**

ITEM	60 INPUT	60 MEASURED		INFIT		OUTFIT	
MEAN	200.0	333.0	.00	.14	1.00	-.1	1.02	.0
S.D.	71.1	.0	1.24	.03	.11	2.0	.22	1.9
REAL RMSE	.15	TRUE SD	1.23	SEPARATION	8.43		**ITEM RELIABILITY**	**.99**

Calculating Fit Statistics

>====================================<

Standardized Residuals N(0,1) **Mean: .00 S.D.: 1.00**

Time for estimation: 0:0:0.169

VELT Version 2

PERSON	288 INPUT	288 MEASURED		INFIT		OUTFIT	
	SCORE	COUNT	MEASURE	ERROR	IMNSQ	ZSTD	OMNSQ	ZSTD
MEAN	37.5	60.0	.81	.34	1.00	.0	1.00	.0
S.D.	9.4	.0	1.04	.07	.14	1.0	.39	.9
REAL RMSE	.35	TRUE SD	.98	SEPARATION	2.79		**PERSON RELIABILITY**	**.89**

ITEM	60 INPUT	60 MEASURED		INFIT		OUTFIT	
MEAN	180.0	288.0	.00	.16	1.00	.0	1.00	.0
S.D.	63.2	.1	1.37	.06	.11	1.7	.23	1.7
REAL RMSE	.17	TRUE SD	1.36	SEPARATION	7.90		**ITEM RELIABILITY**	**.98**

How to contact VIEWS:

VIEWS c/o
Department of English, University of Vienna
Spitalgasse 2-4, Hof 8.3
1090 Wien
AUSTRIA

fax **+ 43 1 4277 9424**
e-mail **views.anglistik@univie.ac.at**
W3 **http://anglistik.univie.ac.at/views/**
(all issues available online)

IMPRESSUM:
EIGENTÜMER, HERAUSGEBER & VERLEGER: VIEWS, C/O INSTITUT FÜR ANGLISTIK & AMERIKANISTIK DER UNIVERSITÄT WIEN, UNIVERSITÄTSCAMPUS AAKH, SPITALGASSE 2, 1090 WIEN, AUSTRIA**. FÜR DEN INHALT VERANTWORTLICH**: ANITA SANTNER-WOLFARTSBERGER, CLAUDIO SCHEKULIN **REDAKTION:** ANDREAS BAUMANN, CHRISTIANE DALTON-PUFFER, NORA DORN, MALGORZATA FABISZAK, HELEN HEANEY, CORNELIA HÜLMBAUER, GUNTHER KALTENBÖCK, KAMIL KAŹMIERSKI, EVELIEN KEIZER, ARNE LOHMANN, NIKOLAUS RITT, BARBARA SCHIFTNER, BARBARA SEIDLHOFER, LOTTE SOMMERER, BARBARA SOUKUP, UTE SMIT, HENRY G. WIDDOWSON, EVA ZEHENTNER. **ALLE:** C/O INSTITUT FÜR ANGLISTIK & AMERIKANISTIK DER UNIVERSITÄT WIEN, UNIVERSITÄTSCAMPUS AAKH, SPITALGASSE 2, A - 1090 WIEN. **HERSTELLUNG:** VIEWS