VIEW[z]

VIENNA ENGLISH WORKING PAPERS

CONTENTS

LETTER FROM THE EDITORS

Dear Readers,

Welcome to the 'hot' summer issue of VIEWS – this time packed with contributions to suit nearly every linguistic taste, from the highly theoretical to the classroom-based researcher.

The first contribution by Ursula Lutzky addresses the wide-ranging and somewhat elusive category of discourse markers, and aims at a clearer delimitation of what features the members of this category actually share and how they might be identified. Apart from bringing some clarity into the matter, ultimately, she also shows which means of classification are relevant in the context of historical linguistics.

Lotte Sommerer tackles a rather theoretical issue in addressing first language acquisition. She presents a network-based model to explain the acquisition of two-word syntax without taking recourse to the assumption of an innate UG, but by explaining all language learning as part of more general learning processes.

Finally, Susanne Sweeney-Novak presents a topical issue for many language teaching departments, namely the move towards standardised assessment procedures. Her paper describes research conducted to accompany the process of implementing new means of standardised assessment – the so-called 'Common Final Test' at Vienna's English Department.

With some chance of a long summer ahead of us, we hope you will find much food for thought in this issue, and maybe even some responses in yourselves that you might wish to send us in the form of written reply – so why not send us a postcard!

<div align="right">

THE EDITORS

</div>

Discourse Markers? Well, … Delimiting the basic features of discourse markers

Ursula Lutzky, Vienna*

> *'But I'm* not *a serpent, I tell you!' said Alice. 'I'm a – I'm a –'*
> *'Well!* What *are you?' said the Pigeon. 'I can see you're trying to invent something!'*
>
> *(Carroll 1994 [1865]: 53)*

1. Introduction

In 1976, Robert E. Longacre referred to a group of 'mystery particles' that had so far been regarded as elements that were "simply salt-and-peppered through a text to give it flavor, … to make it sound like so-and-so language or so-and-so style within that language" (Longacre 1976: 468), i.e. as being in free variation. Longacre was one of the first to notice that the mysteriousness of these particles, which could not be accounted for within sentence grammar, might be resolved when going beyond the level of the sentence. Although the term 'discourse analysis', used for the first time by Zellig Harris in 1952, had been around for some time, it was thus not until the 1970s that the analysis of discourse moved into the centre of linguistic interest and that the nature of discourse was increasingly discussed. It was through the study of "the organization of language above the sentence" (Stubbs 1983: 1)[1] that the mystery of particular discourse phenomena, which had defied analysis within the domain of the sentence, could be unravelled, and what Longacre used to call 'mystery particles' are today regarded as items serving "important

---

* The author can be contacted under ursula.lutzky@univie.ac.at

[1] For a critical discussion of Stubbs 1983 and his definition of discourse analysis see Widdowson 2004.

functions on the textual as well as on the interpersonal levels of spoken discourse" (Lenk 2005: www.benjamins.com/online/hop, 17 May 2006).

Nevertheless, some of their former mysteriousness seems to remain as there is neither a commonly used term to designate this group of 'mystery particles'[2], nor a generally accepted, or even a widely used definition for these expressions. This is, on the one hand, due to the fact that the various approaches to discourse marker[3] research differ as to their theoretical background, while, on the other hand, individual studies based their definitions on different discourse marker functions which they see as primary (cf. Brinton 1996: 30f.). Although the main tendency so far has been to regard discourse markers as items which make the relation between one part of discourse and the preceding discourse or the surrounding context explicit (cf. e.g. Fraser 1988, 1990, 1996, 1999; Lenk 1998a, 1998b; Redeker 1990, 1991; Schiffrin 1987), "not one single definition of the term *discourse marker* remained undisputed or unaltered" (Lenk 1998b: 37). This is also reflected in the unstable inventory of the class of discourse markers: there is no general agreement as to which expressions this class comprises, and so Schiffrin (2001: 65) comes to the conclusion that

> [d]iscourse markers are parts of language that scholars want to study, even if they do not always agree on what particular parts they are studying or what to call the object of their interest.

---

2 Among others, 'mystery particles' have been termed discourse markers (Blakemore 1987, 2002; Fraser 1988, 1990, 1998; Lenk 1998a, 1998b, 2005; Schiffrin 1987), discourse particles (Aijmer 2002; Schourup 1985), discourse operators (Redeker 1990, 1991), pragmatic markers (Andersen 1998, 2001; Brinton 1990, 1995, 1996, 1998; Erman 2001; Fraser 1996), pragmatic particles (Östman 1981, 1982), pragmatic formatives (Fraser 1987), pragmatic expressions (Erman 1986), pragmatic connectives (Crystal and Davy 1975; van Dijk 1979), fillers (Brown and Yule 1983), conjunctions and continuatives (Halliday and Hasan 1976), continuers (Schegloff 1982), cue phrases (Knott and Sanders 1998), or fumbles (Edmondson 1981).

3 While the term 'discourse marker' is used in this article, it is not employed with a restricted range of application (cf. Fraser 1988, 1990, 1996, 1999; Andersen 2001) but as a general cover term. This choice seems justified in so far as the more narrow usage of this term (discourse markers being regarded as a subgroup of pragmatic markers and as signalling the structural organization of discourse only) has not been generally accepted. Furthermore, this term is considered suitable as it indicates that the expressions referred to operate on the level of discourse without, however, being too restrictive as far as their functions are concerned (cf. other labels like 'connectives', 'continuatives').

Consequently, it seems justified to say that "the field of discourse marker research is far from being homogeneous or unified" (Lenk 1998b: 37).

The present article reports some of the findings which form part of a larger project on discourse markers in Early Modern English. In the delimitation of the field of discourse markers, one basically has to take two points into consideration: on the one hand, one needs to account for the pragmatic nature of discourse markers, i.e. for the particular pragmatic functions they may serve. On the other hand, the question of how to identify discourse markers formally in a text needs to be addressed (form-to-function mapping). This article is mainly concerned with the latter question, leaving the discussion of discourse marker functions for a later stage. It attempts to critically discuss the basic features that have been said to characterize discourse markers with a view to their applicability and in particular their practicality in the identification of discourse markers in historical data.

While discourse markers are usually regarded as a feature of oral discourse and research has so far mainly concentrated on spoken Present Day English data, it has been shown more recently that they may also be found in texts from past periods of the English language (cf. Brinton 1990, 1995, 1996, 1998; Jucker 2002; Kryk-Kastovsky 1998; Schmied 1998). The main aim of this paper is to demonstrate that not all of the basic characteristics of discourse markers prove equally useful when one tries to identify them in written historical discourse, while some of them even turn out to be impractical in an analysis of spoken PDE data. It will be argued below that some predominantly formal features have to be regarded as accidental rather than essential, i.e. they may be applied after a discourse marker has been identified but are "not suitable as a test for class-membership" (Jucker and Ziv 1998: 4). Furthermore, it will be shown that particular features are not equally applicable or reliable depending on whether one intends to analyse historical or PDE data. The findings presented in this article will form the basis of a later corpus-based analysis of discourse markers in EModE texts. The identification criteria which are theoretically discussed here will be used in order to distinguish discourse marker from non-discourse marker uses of lexemes.

2. Discourse marker characteristics

As the above discussion suggests, "'discourse marker' is a fuzzy concept" (Jucker and Ziv 1998: 2). Nevertheless, some basic characteristics shared by discourse markers have been identified. Table 1 sums up several of these

features, which were first collected by Brinton (1996: 33-35) and later reordered according to levels of linguistic description by Jucker and Ziv (1998: 3).

- • Phonological and lexical features:
  - a) They are short and phonologically reduced.
  - b) They form a separate tone group.
  - c) They are marginal forms and hence difficult to place within a traditional word class.
- • Syntactic features:
  - d) They are restricted to sentence-initial position
  - e) They occur outside the syntactic structure or they are only loosely attached to it.
  - f) They are optional.
- • Semantic feature:
  - g) They have little or no propositional meaning.
- • Functional feature:
  - h) They are multifunctional, operating on several linguistic levels simultaneously.
- • Sociolinguistic and stylistic features:
  - i) They are a feature of oral rather than written discourse and are associated with informality.
  - j) They appear with high frequency.
  - k) They are stylistically stigmatised.
  - l) They are gender specific and more typical of women's speech.

Table 1: List of basic features of discourse markers (Jucker and Ziv 1998: 3; based on Brinton 1996: 33-35)

While this list comprises features that various studies found to be characteristic of discourse markers, not each form that has been attributed to the class of discourse markers necessarily shows all of them. Consequently, one may distinguish between prototypical and more peripheral members of the class of discourse markers: "[p]rototypical discourse markers will exhibit most or all of these features; less prototypical markers will have fewer features or exhibit them to a limited extent only" (Jucker 2002: 211). In order to account for both – more and less prototypical – types of members, it was thus suggested to view the class of discourse markers as a scale. (Jucker and Ziv 1998: 2ff.; Jucker 2002: 211)

2.1. Phonological and lexical features

As far as phonological features are concerned, discourse markers are said to be phonologically reduced and to form a separate tone group. Obviously, these phonological characteristics seem to be primarily significant for the

identification of discourse markers in studies based on spoken PDE data. As phonetic information is lacking for past periods of the English language, they appear to be of little use to an analysis which aims at identifying discourse markers in early English texts. However, Brinton (1995: 379) shows that – to some extent – the feature "phonetic 'shortness' or reduction" is also applicable to Old and Middle English data: "OE *Þa* exists alongside a longer form *Þanne* or *Þonne*, … while ME *gan* is an aphetic form of OE *on-/aginnan* and occurs concurrently with the prefixed form *beginnen…*". Similarly, Jucker (2002: 211f.) argues that EModE *pray* and *faith* are abbreviations of *I pray you* and *in faith*. Consequently, even though spoken data are not available for past periods of the English language, this phonological feature of discourse markers also seems to be applicable in the analysis of early English texts. However, as Jucker (2002: 212) notes,

> [i]n historical data this criterion is more difficult to apply since there is no way of checking the actual pronunciation beyond the orthographic representation, which may or may not reflect the phonological reduction.

Similarly, also the second phonological feature (discourse markers forming separate tone units) can be taken into account in a study based on written data. This is due to the fact that a tone unit boundary may be signalled by punctuation and a discourse marker may thus be set apart from a following utterance by punctuation marks. This is illustrated by Fraser (1990: 388) in the following examples:

(1) A: John left. *Now*, Mary was really frightened.

(2) A: John left. Now Mary was really frightened.

According to Fraser, *now* is used as a discourse marker in (1) but is said to function as a preposed time adverbial in (2). While Fraser's clear-cut examples illustrate that punctuation may function as a means of distinguishing between discourse marker and non-discourse marker uses of lexemes, it is a means that is not always present. As not all discourse markers are separated from the following utterance by punctuation marks in writing (or form separate tone groups in speaking), ambiguity may arise. Consequently, in an example like "Now where are we?" (Fraser 1996: 170), where there is no comma (intonation), *now* is ambiguous between an adverb of time and a discourse marker.

As far as the analysis of early English texts is concerned, it is important to stress that Modern English punctuation differs considerably from

punctuation in earlier periods of the English language. In an analysis of a text sample of Ælfric's *Catholic Homilies,* Blake (1996: 101) notes that

> *[t]he punctuation seems to be erratic, for there is little to indicate why particular marks occur where they do … The marks also occur at regular intervals and thus appear to indicate pauses for breathing or for rhetorical effects rather than for clarification of the grammatical arrangement of the text.*

Consequently, punctuation, which is used to fulfil a rhetorical rather than a grammatical function, is not a very reliable criterion before its standardisation in EModE. This is also because different editions of one and the same text differ in their punctuation, which suggests that punctuation was added or changed by editors (Blake 1996: 161f., 241f.; Jucker 2002: 212f.). Thus, in her analysis of OE and ME text samples Brinton (1996: 266) can only conclude that "modern editing of texts would suggest that at least certain forms … should be viewed as separate intonational units".

Furthermore, discourse markers are said to be short expressions and the inventory generally ranges from single lexical items to two- or three-word phrases. However, individual studies also subsumed various longer and even clausal expressions under the category of discourse markers (cf. for example Fraser 1996 who includes the following examples in his discourse marker class: *on the other hand*, *by the same token*, *on a different note*, *on top of it all*, *it stands to reason that*, *it can be concluded that*). The questions that arise are, on the one hand, whether one should regard these expressions as short – especially in comparison to discourse markers like *well* or *now* – and, on the other hand, whether one should actually include them in the class of discourse markers. The fact that discourse markers are said to be "a feature of oral rather than written discourse and are associated with informality" (Jucker and Ziv 1998: 4) would speak against regarding the examples cited above as discourse markers, as they would rather be associated with written and formal discourse. Since they also fail to exhibit many of the other basic features mentioned in Table 1 (phonological, semantic, syntactic, stylistic/sociolinguistic), it is questionable if they should be regarded as peripheral members of the discourse marker group or if one should attribute them to a separate category of expressions serving similar functions to discourse markers but not sharing all of their other characteristics. For example, Lenk (1998b: 50) argues that "not every lexical item expressing e.g. juxtaposition or coordination of different elements automatically classifies as a discourse marker". On these grounds, she excludes lexical expressions that have a structuring function, like *to return to my point, as a result, in other*

words[4], from the class of discourse markers as they only serve this one function but are not multifunctional like discourse markers proper.

Apart from being lexically and phonetically short, discourse markers are also said to be 'marginal' forms. They are a heterogeneous set of expressions which do not stem from a single grammatical category but from various sources:

> *adverbials (*now*, then*, still*), literally used phrases (*to repeat*, what I mean to say*, similarly*, overall*), idiomatic phrases (*while I have you*, still and all*), verbs (*look*, see*), interjections (*well*), coordinate conjunctions (*and*, or*, but*), subordinate conjunctions (*so*, however*), as well as terms such as* anyway *and* OK*, which don't fall nicely into any of the usual grammatical slots. In short, discourse markers are not adverbs, for example, masquerading as another category from time to time. (Fraser 1990: 388)*

Consequently, a classification according to traditional word classes seems to be out of the question, though there have been attempts at classifying discourse markers as adverbs – as the above quote implies – but also as particles, conjunctions or interjections (cf. e.g. Stubbs 1983, Halliday and Hasan 1976). A lexical classification is further complicated by the fact that with some items which may function as discourse markers there is not even agreement as to which class they belong to (cf. *oh* which is classed as an interjection and as an exclamation in the *OED*; or *well* which according to Fraser forms part of the class of interjections while one may also attribute it to the category of adverbials).

In short, discourse markers are, on the one hand, formally very diverse; on the other hand, they are functionally similar. While the notion of a 'class' of discourse markers has been questioned due to their formal diversity, attempts have been made to base group membership on their "various degrees of functional similarities and partially overlapping distributions" (Schiffrin 1987: 65).

Jucker and Ziv (1998: 4) include the phonological and lexical features among the diagnostic features of discourse markers – those "which provide the crucial tests". While I agree that these features are characteristic of prototypical discourse markers, it should, however, be clear that they are not sufficient as defining features but need to be complemented by additional ones. Nevertheless, the phonological and lexical characteristics (with the exception of lexical shortness) can be regarded as practical criteria by which one may distinguish discourse marker from non-discourse marker uses of

---

[4] These are examples that Fraser (1996: 186ff.) includes in his class of discourse markers.

lexemes. As far as historical data are concerned, however, it has been shown that the phonological features – while applicable – do not prove very reliable.

2.2. Syntactic features

The syntactic characteristics of discourse markers mentioned in Table 1 imply that discourse markers occur at the level of the sentence. However, the notion of 'sentence' may be considered inappropriate with regard to the occurrence of discourse markers in both spoken and written data. While discourse markers were said to occur at boundaries, these cannot be regarded as sentence boundaries if we consider Widdowson's distinction between 'sentence' and 'utterance':

> *A sentence has only one invariant meaning, or if it has more than one, as in the case of structural or lexical ambiguity, its meanings can be exactly specified. Utterances, on the other hand, are protean in character. Their meanings change continually to suit the circumstances in which they are used. (Widdowson 1990: 100)*

Consequently, one may rather say that discourse markers occur at the level of the utterance –an argument that is also reflected in Schiffrin (1987: 35), who in her definition of discourse markers comes to the "deliberately vague conclusion that markers bracket units of talk".

Coming back to the syntactic characteristics of discourse markers listed by Jucker and Ziv, their first syntactic feature defines them as being "restricted to sentence-initial position" (Jucker and Ziv 1998: 3). However, discourse markers, though typically appearing initially in a unit of talk, are not restricted to this position but may also be found in medial and final position[5]. In fact, it could be shown that "many discourse markers are flexible and can appear in different positions in the utterance" (Lenk 2005: www.benjamins.com/online/hop, 17 May 2006) and it was even claimed that "certain markers can seemingly appear virtually anywhere within an utterance" (Andersen 2001: 48). When used in medial or final position, discourse markers are, however, usually set off by a comma (intonation) in order to distinguish them from homophonous forms expressing propositional meaning. Thus, apart from position, punctuation and intonation (i.e. discourse markers' parenthetical nature) are important clues which may indicate whether particular expressions are to be interpreted as carrying propositional

---

[5] Brinton (1996: 33), on which Jucker and Ziv's list of discourse markers' basic features is based, explicitly states that pragmatic markers also frequently appear in medial and final position.

meaning or fulfilling discourse marker functions. As these clues are, however, not always present and in the case of historical data not reliable themselves, ambiguity between sentential and discourse uses of a lexeme may arise and misunderstandings or confusion may result (cf. 2.1 above). (Fraser 1988: 24, 1990: 389, 1996: 170; Lenk 1998b: 51f.; Schiffrin 1987: 328)

The second syntactic criterion describes discourse markers as occurring outside the syntactic structure or being only loosely attached to it. Brinton (1996: 34) adds that they "hence have no clear grammatical function". Compared to position and the phonological and lexical characteristics of discourse markers discussed above, this appears to be a more reliable identifying feature of discourse markers that is applicable to both spoken and written data and useful when it comes to distinguishing discourse marker from non-discourse marker uses of a form. For example, the lexeme *now* may appear both as an adverbial and as a discourse marker in an utterance. While the adverb forms part of the syntactic structure of an utterance (time-adjunct) and may be highlighted in a cleft sentence, discourse marker *now* does not serve a grammatical function in an utterance and cannot be cleft-highlighted. Apart from that, Kryk-Kastovsky (1995: 88) mentions that pragmatic particles "do not form structural constituents with other lexical items, i.e. … they are immune to modifications". Thus the adverb and the discourse marker uses of *well* may be distinguished in so far as the former can be modified (e.g. *very well*) while the latter cannot. Furthermore, as Fischer (In press: chapter 6) points out, pragmatic markers do not cause inversion when occurring in initial position in OE and still do not do so in modern Germanic languages such as Dutch or German. In these Verb-second languages, pragmatic markers (cf. *hwæt* in OE) do not entail inversion of subject and verb but are, as Fischer demonstrates, followed by SV/XV order. This indicates that they do not belong to the syntactic structure of the following utterance but have to be regarded as separate phrases. On the other hand, adverbials which appear in initial position and are followed by VS/VX order were found not to constitute separate phrases but to form part of the matrix clause.

This last syntactic characteristic is closely linked to the remaining syntactic feature of discourse markers mentioned in Table 1: the fact that discourse markers are said to occur outside the syntactic structure implies that they are optional (cf. Fraser 1988, who calls them 'lexical adjuncts'). Consequently, they can normally be left out without rendering the discourse either ungrammatical or unintelligible. This is also due to the assumption that discourse markers "do not create meaning relationships in texts but merely reveal or make explicit those connections already operating in texts" (Brinton

1996: 267; cf. Fraser 1990: 390) and is supported by the fact that discourse markers are not constrained to a particular position in an utterance but show "great syntactic freedom" (Andersen 2001: 48; cf. also Brinton 1996: 34f., 267; Fraser 1988: 22).

While syntactic independence and the resulting optionality are important defining features of discourse markers, I do not agree with Müller (2005: 6) who regards grammatical optionality as

> *the only [feature] … which can be used to distinguish, for example, between discourse markers and their non-discourse marker homonyms (e.g.* well *as an adverb,* you know *in questions).*

On the contrary, I would rather say that this characteristic cannot always be applied as a distinguishing feature between discourse marker and non-discourse marker uses of a lexeme. Adjuncts, for instance, are usually omissible without rendering an utterance ungrammatical or unintelligible (cf. Müller's example: "*well* as an adverb"). What is changed by their omission is the propositional content of an utterance, but it has no impact on its grammatical well-formedness. Since the feature of being syntactically optional thus does not apply to discourse markers alone but may also characterize their non-discourse marker homonyms, I would refrain from calling it the *only* distinguishing feature between these two uses.

It is, however, important to point out that while discourse markers do not have a grammatical function and their omission consequently does not affect the grammaticality of an utterance, they do serve important pragmatic functions. Without discourse markers, transitions between turns may seem disjointed or abrupt and ambiguities may arise more easily, as what is removed is "a powerful clue about what commitment the speaker makes regarding the relationship between the current utterance and the prior discourse" (Fraser 1988: 22). If discourse markers are omitted, the relationship between two parts of discourse is no longer explicitly signalled by lexical means. Consequently, the hearer has to rely on implicit clues (e.g. intonation, context) and as a result communication may break down more easily (cf. Fraser 1990: 390, 1996: 186f., 1999: 944). As there are, however, other devices (e.g. repetition, thematic continuity…) that may fulfil more or less the same functions as discourse markers, the contribution of a discourse marker to the overall structure of a discourse as well as the necessity for additional markers to occur may be reduced. Schiffrin (1987: 322) thus concludes that

the more the discourse works toward conveying its own meaning and structure, the smaller the contribution of the discourse marker, and the more the marker is likely to be absent.

2.3. Semantic features

Discourse markers are usually said to carry little or no propositional meaning and thus do not contribute to the propositional content of an utterance. Therefore, discourse markers do not change the truth value of an utterance, i.e. "a true sentence is true, and a false sentence is false, whether or not it contains a discourse marker" (Jucker 2002: 213). As it is consequently rather difficult to specify the meaning of a discourse marker, they generally pose a problem in translation. According to Brinton (1995: 379), this is why many of the OE and ME markers used to be regarded as mystery features and were "traditionally described as 'meaningless', 'empty', or 'colorless'".

While this feature is a widely accepted characteristic of discourse markers (Brinton 1990, 1995, 1996, 1998; Crystal and Davy 1975; Edmondson 1981; Jucker 1993, 2002; Jucker and Ziv 1998; Levinson 1983; Östman 1981, 1982; Schiffrin 1987; Stubbs 1983; Svartvik 1980), Andersen questions this assumption and states that non-propositionality is not "an essential property of pragmatic markers" (Andersen 2001: 40). He argues that only some pragmatic markers can be readily classified as non-propositional (e.g. *ah*, *and*, *oh*, *moreover*, *so*, *well*, *uh huh* …), whereas others may also have truth-conditional implications (e.g. *you know*, *I mean*) and should be regarded as moving along a continuum with propositional and non-propositional uses as its end-points. The reason why some of the expressions which originated in lexical forms with propositionally encoded meanings cannot be easily classified according to the propositional/non-propositional dichotomy has to do with their diachronic development and their degree of grammaticalisation. According to Andersen, the grammaticalisation process has not yet been completed, i.e. "resulted in clear polysemous forms" (Andersen 2001: 52), in the case of those pragmatic markers which cannot be readily classified as non-propositional. (Andersen 2001: 38ff.)

Based on these assumptions, Andersen (2001: 57f.) devised a diachronic model in which the development from propositional lexeme to pragmatic marker is presented as a three-stage process: first, a propositional expression can be described as monosemous. Then it adopts "new functions and more opaque meanings than the original" and this innovative use "gradually becomes conventionalised" (Andersen 2001: 57). It is at this intermediate stage that the grammaticalisation process is in progress and that expressions

move along the propositional/non-propositional continuum. The final stage is then the polysemous state which "is characterised by greater fixedness and distinctness of the two functions, as the invited inference that was firstly innovative has become routinised and part of the linguistic code" (Andersen 2001: 57).[6] Andersen thus comes to the conclusion that

> *despite the common correlation of markerhood and procedural encoding, we cannot rule out that some pragmatic markers may be conceptual. After all, pragmatic markers constitute a broad category that [also] includes … multi-word items like I mean, you know, I think and sort of. … Therefore, the conceptual/procedural distinction cannot be applied as a definitional criterion to characterise the pragmatic marker category. (Andersen 2001: 61f.)[7]*

While I agree with Andersen as far as the diachronic development of some discourse markers from propositional lexemes is concerned[8], I would, nevertheless, propose that non-propositionality should not be discarded as an essential, defining feature of discourse markers. Together with the two syntactic features of optionality and occurrence outside the core syntactic structure, I would rather place non-propositionality among the main defining features of discourse markers. It is this semantic characteristic which constitutes an important distinguishing feature between pragmatic and non-pragmatic functions of a lexeme (e.g. *well* used as an adverb contributes to the propositional content of an utterance, while discourse marker *well* does not). Though it may not be sufficient as a defining feature of discourse markers, it is one of the most practical in a test for class-membership. Contrary to Andersen, I would thus argue against the exclusion of non-propositionality as a definitional criterion of discourse markers and, adopting Jucker and Ziv's (1998) prototypical approach, rather view expressions that have not reached the polysemous state as less prototypical members of the class. According to this approach, more prototypical discourse markers such as *well* would be said to satisfy the semantic criterion of non-propositionality, "while the more marginal discourse markers *y'know* and *like* have some residue of semantic meaning" (Jucker 2002: 213).

---

6 Andersen (2001: 57) points out that an expression need not complete this three-stage development and reach the polysemous state. On the other hand, expressions that do reach this state may develop further, i.e. this state is not to be regarded as final.

7 Andersen 2001, working within a relevance theoretic framework, uses the terms 'conceptual' and 'procedural' to refer to what I call 'propositional' and 'non-propositional'.

8 As Andersen (2001: 58) points out, it is important to bear in mind that not all discourse markers have developed from propositional lexemes, cf. e.g. *ah*, *oh*, or *uh huh*.

2.4. Functional features

Discourse markers are usually said to be multifunctional. While this is a widely accepted characteristic of discourse markers, one needs to distinguish between two types of multifunctionality: on the one hand, expressions which may be used as discourse markers may be regarded as multifunctional because they can also have non-pragmatic functions in a text. For example, the lexeme *well* can, apart from its discourse marker uses, function as the head of a noun, verb, or adverb phrase. On the other hand, one may consider multifunctionality within the domain of pragmatics. That is to say that one and the same discourse marker may serve various pragmatic functions at different levels.

Some studies restrict the term multifunctional to the pragmatic domain: discourse markers are regarded as multifunctional because "they can serve different pragmatic functions in different contexts" and may also display "several pragmatic features at the same time" (Andersen 2001: 64). While Andersen takes account of grammaticalisation and the possible development of an original lexical item into a grammaticalised pragmatic marker (cf. above), Östman (1982: 153) takes a more radical approach and states that "for an item to be called a pragmatic particle, it should never be able to have any other than a pragmatic-particle function". Although Östman speaks of a 'sole function' of pragmatic particles in relation to his 'uniqueness criterion', he nevertheless assumes that they may serve several pragmatic sub-functions.

On the other hand, other scholars (cf. e.g. Brinton 1996, 1998; Lenk 1998; Redeker 1990; Schiffrin 1987) do not limit multifunctionality to the pragmatic domain but take account of both types of multifunctionality in their discourse marker studies. Schiffrin (1987), for instance, regards discourse markers as functioning on various planes of talk (exchange, action, ideational structure; participation framework; information state) and notes that they may operate on more than one plane simultaneously. She defines the ideational structure as semantic and claims that "the units within this structure are … propositions, or what I'll just call ideas" (Schiffrin 1987: 25), and thus assumes that discourse markers may operate on the semantic and the pragmatic level at the same time. While one may, according to Schiffrin, distinguish between primary and secondary planes of use of a particular marker, it is not always obvious which function is predominant in a particular context. This implies that the pragmatic and non-pragmatic uses of a form may not always be easily distinguishable. This assumption is, however, in contrast to Östman's statement that "the pragmatic and propositional

functions [of an expression] are clearly separate in nature, with no scalar relation between the two" (Östman 1982: 153f.)

Similarly to Östman, Fraser claims that "when an expression functions as a discourse marker, that is its exclusive function in the sentence" (Fraser 1990: 389), even though it may have homophonous forms serving different functions (e.g. *now* as an adverbial). He continues to say that "there is never a doubt when an expression is functioning as a discourse marker" (Fraser 1998: 257)[9]. This argumentation can, however, be easily refuted, as it is by no means always clear whether a particular form serves pragmatic functions or carries propositional meaning. While the different functions of the lexeme *however* are clearly recognisable in the (presumably invented) example quoted in Fraser (1998: 257),

> (3) John wants to take him. However, he will have to get there however he can.

more ambiguous examples have been identified. Consider, for instance, the following example, where the distinction between what Lenk (1998b: 101ff.) regards as the propositional and the discourse marker functions of *however* is blurred:

> a [@m] ^does – operational re:search# is it is it . ^primarily concerned# with ^questions of {distribution} – marketing#

> b indeed no# - - - ^operation research started# - - - ^by studies# - of ^military problems# - - - ^one of the first# ^exercises ever carried out# - - took ^place during the war# - - - ^when - - the question# . of ^whether small boats# . should ^carry anti-:aircraft guns# - was con^sidered# - - ^these small boats had been equipped# - - with ^anti-aircraft guns# - - but they ^weren't shooting down# - ^any more enemy aircraft# - - and ^therefore# ^certain people concluded# - that ^these guns# . ^weren't fulfilling their function# - - - **how^\ever#** ^when the operation research man looked# - - at the ^data more closely# - he dis^covered# . that ^fewer boats were being sunk# - in ^other words#

> *(LLC 6.1.1190-1217; quoted in Lenk 1998b: 109)*

According to Lenk, *however* in this example could express contrast but it could also mark "the end of a short digression which was of considerable importance to the development of the topic" (Lenk 1998b: 109), and only in the latter case it would serve a discourse marker function. In contrast to

---

9 Note that in a previous article Fraser (1990: 170), however, states that ambiguity may arise in a few examples, where the pragmatic and non-pragmatic uses of a lexeme cannot be easily kept apart (cf. 2.1. above).

Fraser, Lenk (1998: 105ff.) thus does not regard *however* as a discourse marker when it signals contrast but lists this function among the propositional uses of *however*. She concludes that in the case of *however* the propositional and discourse marker functions cannot always be easily kept apart but are closely related, as the above example illustrates.[10]

Lenk's example also shows that propositional and non-propositional occurrences of a lexeme cannot generally be said, as Fraser (1998, 1999) argues, to be in syntactically complementary distribution. One suspects that Fraser's claim may apply only to his constructed example sentences (cf. Fraser 1999: 944, "*He didn't want to go. On the other hand, he didn't want to stay.*" vs. "*One hand was unadorned. He had a colourful tattoo on the other hand.*").

Concerning the practicality of this basic feature of discourse markers, I do not include multifunctionality among the group of decisive characteristics as it does not constitute a relevant identification criterion of discourse markers in either spoken or written discourse. As Jucker and Ziv (1998: 4) argue

> *[w]hether a specific linguistic element is monofunctional or polyfunctional is not a useful criterion in deciding whether it is a discourse marker or not because of the obvious analytical vicious circularity it entails.*

2.5. Sociolinguistic and stylistic features

Similarly to the functional feature, the sociolinguistic and stylistic features must also be regarded as primarily descriptive and accidental. They are of no use for the identification of discourse markers but, as Jucker and Ziv (1998: 4) argue, "the sociolinguistic and stylistic distribution can only be established once a discourse marker has been identified as such". The following discussion will consequently not be concerned with the practicality of these features as a test for class-membership. Instead, the basic applicability of these features to both historical and PDE data will be discussed and areas for further research will be pointed out.

First, discourse markers are said to be primarily a feature of oral rather than written discourse. Consequently, one may assume that they are less likely to be attested in historical data and that historical discourse marker research may prove impossible. However, as Brinton (1995: 377) notes

---

10 This also shows how various studies may differ in their definition of discourse markers and consequently also in what they regard as discourse marker functions.

> *[w]hile it may initially seem implausible to speak of pragmatic markers in texts which exist only in written form, it is generally agreed that medieval discourse exhibits many structural and linguistic elements characteristic of oral discourse (Brinton 1995: 377).*

Apart from medieval discourse, which is said to contain an "oral residue" (Ong 1984) as it marks the transition from an oral to a literate culture, genres that "represent spoken language in a written form" (Jucker 2002: 210) were also shown to contain features of orality. In an analysis of discourse markers in EModE, Jucker (2002: 213) discovered that "the frequency of discourse markers is a direct correlate of the amount of (representations of) spoken language that is likely to occur in any particular genre". Thus, discourse markers were attested with considerable frequencies in EModE text types that are related to spoken language, i.e. records of spoken language (e.g. court records), texts written to be spoken (e.g. sermons) or texts imitating spoken language (e.g. plays).

While discourse markers are generally said to be a feature of oral rather than written discourse, this does not mean that "[t]hey are all restricted to spoken language" (Erman 2001: 1339). On the one hand, this is confirmed by historical studies on discourse markers (cf. above), whereas, on the other hand, discourse markers were also identified in PDE written discourse that "has a high degree of impromptuness" (Östman 1982: 170). As far as PDE data are concerned, however, discourse marker research has so far mainly concentrated on spoken data and a large-scale analysis of discourse markers in PDE written texts is, in fact, still missing. It is this kind of study that would, however, provide empirical evidence for Fraser's (1990: 389) intuitive statement "that certain discourse markers occur more frequently in written discourse (e.g. *notwithstanding*) while others are found more frequently in conversation (e.g. *OK*)". Furthermore, the analysis of discourse markers in written discourse may also reveal whether particular expressions, like *notwithstanding*, should indeed be included in the class of discourse markers or whether one should rather attribute them to a separate group of markers serving similar functions to discourse markers but appearing primarily in written discourse.

Apart from the prevalence of discourse markers in oral discourse, they are also said to be associated with informality (cf. Table1, i). On the one hand, this feature is linked to the primarily oral nature of discourse markers and Brinton, for example, states that "[t]he appearance of pragmatic markers is a result of the informality of oral discourse" (Brinton 1996: 33, cf. Östman 1982). On the other hand, oral discourse should not generally be regarded as

informal. Thus, Aijmer (2002: 34) argues that although "a correlation between informal conversation and discourse particles" has been proposed by several scholars (cf. Biber 1988, Östman 1982), the occurrence of individual particles may differ as far as register is concerned. In an analysis of discourse particles in the London-Lund Corpus of Spoken English, she was able to show that they differ in terms of text type distribution. For example, "[s]ort of is symptomatic of informal speech" (Aijmer 2002: 190), whereas *now* was found to appear primarily in text types that are "more formal than ordinary conversation and contain more structure" (Aijmer 2002: 69). In Jucker and Ziv's prototype framework the feature of informality is regarded as being primarily characteristic of more prototypical discourse markers whereas more peripheral members need not satisfy this criterion. However, Aijmer's text type analysis shows that a more detailed study of the distribution of discourse markers across different text types may reveal interesting results as far as the formal – informal gradation is concerned.

A further characteristic of discourse markers is their "stylistic stigmatization". Regarding this feature two questions arise: by whom and in what context have discourse markers been stylistically deplored?   As Brinton (1996: 33) notes, it is due to their high frequency as well as their oral nature that discourse markers have been especially stigmatized in written and formal discourse. Also, their apparent lack of meaning contributed to the fact that their occurrence in OE and ME texts used to be deplored. Concerning the lexeme *gan,* Brinton (1995: 379) thus states that

> [w]hen seen as a metrical expedient, inserted by (incompetent) poets to add a syllable to a line of verse or to move the infinitive into rhyme position, gan *is considered a defect of ME style.*

However, such stylistic stigmatization may also reflect that ModE standards were wrongly applied to past periods of the English language. Thus Brinton (1998: 10) mentions that as the phrase *then it happened that* (*Ða gelamp hit Þæt*) and the parenthetical *I guess* (*I gesse*) are mainly used in oral and colloquial discourse in ModE, they were also considered stylistically inappropriate in OE and ME texts, even though these expressions were attested in various types of texts written by different authors. It was thus by speakers of ModE and in formal and literary texts that the use of these colloquial expressions was negatively evaluated.

However, discourse markers were stylistically stigmatized not only in early English texts but also in ModE their use has been "deplored as a sign of dysfluency and carelessness" (Brinton 1996: 33). Thus, for example, Watts (1989) discovered in a study of the discourse markers *you know*, *right* and

well that native speakers make (sometimes frequent) use of them but do not seem to be aware of it as they condemn their numerous attestations in other people's speech. Interestingly, though, the studies from which Brinton derived this characteristic of discourse markers were all conducted in the 1970s and '80s[11]. Despite their findings, the evaluation of discourse marker usage might have changed during the last decades and studies based on speakers' perception of discourse markers today may produce different results. Apart from that, the analysis of speakers' evaluation of discourse marker usage may reveal whether individual markers differ with regard to their stylistic perception. Thus, while the discourse markers *you know*, *I mean*, *I guess* and *like* were said to be negatively evaluated (cf. Östman 1982: 171; Schiffrin 1987: 310f.; Schourup 1985: 39, 94), other discourse markers (e.g. *now*, *so*) may be perceived differently. The integration of sociolinguistic information may, furthermore, disclose in which contexts or text types (e.g. formal – informal, spoken – written etc.) and by whom (e.g. age group, gender etc.) individual markers are in fact stylistically stigmatized.

Finally, the last feature mentioned by Brinton (1996) and Jucker and Ziv (1998) describes discourse markers as being gender specific and more typical of women's speech. Brinton calls this feature "a controversial suggestion" as different studies arrived at different results (cf. Östman 1982, Holmes 1986 for *you know*) and it consequently lacks empirical foundation. Similar to the other sociolinguistic and stylistic features of discourse markers, gender is, moreover, not a useful criterion in deciding whether a particular expression is a discourse marker or not. As Jucker and Ziv (1998: 4) put it,

> *it is unlikely that we would want to exclude a particular element from the set of discourse markers if it turned out that it was gender specific or that it was more common in men's speech.*

Even if this characteristic should turn out to be primarily descriptive, large-scale analyses would still be needed in order to show whether it is indeed justified to list it among discourse markers' basic features at all.

Moreover, as far as the analysis of historical data is concerned, this feature may prove difficult to investigate. This is due to the fact that the majority of texts that have survived from the OE, ME and EModE periods were written by male authors and texts written by women are either not available at all for particular periods or their number is comparatively small. However, we also

---

[11] Brinton (1996: 32ff.) derived her whole inventory of pragmatic marker characteristics from "general studies of pragmatic markers as well as from studies of individual forms".

find records or imitations of female speech in early English texts (cf. court records, plays). While it is debatable how closely these text types resemble actual female language, they might nevertheless provide a starting point for an analysis of the gender-specific use of discourse markers in the history of English, which has not been carried out so far.

3. Conclusion

Summing up, two important points seem to emerge. On the one hand, the majority of features said to characterise discourse markers are descriptive in nature. That is to say that "[n]oncompliance with one of them will rarely lead to an exclusion of the linguistic item from the group of discourse markers" (Müller 2005: 4). On the other hand, no single discourse marker will necessarily show all of the basic features listed in Table 1. Rather, one may distinguish between more and less prototypical discourse markers which differ in the number of basic characteristics they share. (Jucker and Ziv 1998: 4; Jucker 2002: 211; Brinton 1996: 29)

Furthermore, the individual characteristics that have been identified for discourse markers are closely related (cf. the syntactic features 'occurrence outside the syntactic structure' and 'optionality' or the syntactic feature 'position' and the phonological feature 'occurrence in a separate tone unit'). Consequently, while one may arrange the basic features of discourse markers according to levels of linguistic description, they should not be viewed as unconnected, but as overlapping to a certain degree.

As far as the usefulness of the basic features of discourse markers is concerned, not all of them are equally reliable in an analysis of spoken and written discourse. Instead, for the identification of discourse markers in both PDE and historical discourse three criteria can be singled out as primary defining features: occurrence outside of the syntactic structure, optionality and non-propositionality. Even though these are regarded as the most useful criteria for the identification of discourse markers, it can nevertheless be claimed that "[f]ormally and structurally there seems to be no single property or set of properties univocally defining this class" (Aijmer 2002: 27f.). While it is possible to identify individual discourse markers in spoken and written discourse due to their formal properties, a definition of discourse markers cannot be exclusively based on their formal features but needs to take their functions into account.

References

Aijmer, Karin. 2002. *English discourse particles. Evidence from a corpus*. Amsterdam: Benjamins.

Andersen, Gisle. 1998. "The pragmatic marker *like* from a relevance-theoretic perspective". In Jucker, Andreas H.; Ziv, Yael (eds.). *Discourse markers. Descriptions and theory*. Amsterdam: Benjamins, 147-170.

Andersen, Gisle. 2001. *Pragmatic markers and sociolinguistic variation: a relevance theoretic approach to the language of adolescents*. Amsterdam: Benjamins.

Biber, Douglas. 1988. *Variation across speech and writing*. Cambridge: CUP.

Blake, Norman F. 1996. *A history of the English language*. Basingstoke: Macmillan.

Blakemore, Diane. 1987. "Are apposition markers discourse markers?". *Journal of Linguistics* 32, 325-347.

Blakemore, Diane. 2002. *Relevance and linguistic meaning. The semantics and pragmatics of discourse markers*. Cambridge: CUP.

Brinton, Laurel J. 1990. "The development of discourse markers in English". In Fisiak, Jacek (ed.). *Historical linguistics and philology*. Berlin: de Gruyter, 45-71.

Brinton, Laurel J. 1995. "Pragmatic markers in a diachronic perspective". In Ahlers, Jocelyn et al. (eds.). *Proceedings of the twenty-first annual meeting of the Berkeley Linguistics Society. February 17-20, 1995. General session and parasession on historical issues in sociolinguistics / social issues in historical linguistics*. Berkeley: Berkeley Linguistics Society, 377-388.

Brinton, Laurel. 1996. *Pragmatic markers in English. Grammaticalization and discourse functions*. Berlin: de Gruyter.

Brinton, Laurel J. 1998. "'The flowers are lovely; only, they have no scent': The evolution of a pragmatic marker". In Borgmeier, Raimund et al. (eds.). *Anglistentag 1997 Giessen. Proceedings*. Trier: WVT, 9-33.

Brown, Gillian; Yule, George. 1983. *Discourse analysis*. Cambridge: CUP.

Carroll, Lewis. 1865 [1994]. *Alice's adventures in wonderland*. London: Puffin Books.

Crystal, David; Davy, Derek. 1975. *Advanced conversational English*. London: Longman.

Culpeper, Jonathan; Kytö, Merja. 1999a. "Modifying pragmatic force. Hedges in Early Modern English dialogues". In Jucker, Andreas H.; Fritz, Gerd; Lebsanft, Franz (eds.). *Historical dialogue analysis*. Amsterdam: Benjamins, 293-312.

Culpeper, Jonathan; Kytö, Merja. 1999b. "Investigating nonstandard language in a corpus of Early Modern English dialogues". In Taavitsainen, Irma; Melchers, Gunnel; Pahta, Päivi (eds.). *Writing in nonstandard English*. Amsterdam: Benjamins, 171-187.

Edmondson, Willis. 1981. *Spoken discourse. A model for analysis*. London: Longman.

Erman, Britt. 1986. "Some pragmatic expressions in English conversation". In Tottie, Gunnel and Ingegerd Bäcklund (eds.). *English in speech and writing. A symposium*. Stockholm: Almqvist & Wiksell, 131-147.

Erman, Britt. 2001. "Pragmatic markers revisited with a focus on *you know* in adult and adolescent talk". *Journal of Pragmatics* 33, 1337-1359.

Fischer, Olga. In press. *Morphosyntactic change. Functional and formal perspectives*. Oxford: OUP.

Foolen, Ad. 2005. "Pragmatic particles". In Östman, Jan-Ola; Verschueren, Jef (eds.). *Handbook of Pragmatics Online*. Amsterdam: Benjamins.

Fraser, Bruce. 1987. "Pragmatic formatives". In Verschueren, Jef; Bertucelli-Papi, Marcella (eds.). *The pragmatic perspective. Selected papers from the 1985 International Pragmatics Conference*. Amsterdam: Benjamins, 179-194.

Fraser, Bruce. 1988. "Types of English discourse markers". *Acta Linguistica Hungarica* 38(1-4), 19-33.

Fraser, Bruce. 1990. "An approach to discourse markers". *Journal of Pragmatics* 14, 383-395.

Fraser, Bruce. 1996. "Pragmatic markers". *Pragmatics* 6(2), 167-190.

Fraser, Bruce. 1998. "Discourse markers". In Mey, Jacob L. (ed.). *Concise encyclopedia of pragmatics*. Amsterdam: Elsevier, 256-257.

Fraser, Bruce. 1999. "What are discourse markers?". *Journal of Pragmatics* 31, 931-952.

Halliday, M. A. K.; Hasan, Ruqaiya. 1976. *Cohesion in English*. London: Longman.

Harris, Zellig. 1952. "Discourse analysis". *Language* 28, 1-30.

Holmes, Janet. 1986. "Functions of *you know* in women's and men's speech". *Language in Society* 15, 1-22.

Jucker, Andreas H. 1993. "The discourse marker *well*: a relevance-theoretical account". *Journal of Pragmatics* 19, 435-452.

Jucker, Andreas H.; Ziv, Yael. 1998. "Discourse markers: an introduction". In Jucker, Andreas H.; Ziv, Yael (eds.). *Discourse markers. Descriptions and theory*. Amsterdam: Benjamins, 1-12.

Jucker, Andreas H. 2002. "Discourse markers in Early Modern English". In Watts, Richard; Trudgill, Peter (eds.). *Alternative histories of English*. London: Routledge, 210-230.

Knott, Alistair; Sanders, Ted. 1998. "The classification of coherence relations and their linguistic markers: An exploration of two languages". *Journal of Pragmatics* 30, 135-175.

Kryk-Kastovsky, Barbara. 1995. "Particles revisited". In: Melchers, Gunnel; Warren, Beatrice (eds.). *Studies in Anglistics*. Stockholm: Almqvist and Wiksell International, 81-94.

Kryk-Kastovsky, Barbara. 1998. "Pragmatic particles in Early Modern English court trials". In Borgmeier, Raimund et al. (eds.). *Anglistentag 1997 Giessen. Proceedings*. Trier: WVT, 47-56.

Lenk, Uta. 1998a. "Discourse markers and global coherence in conversation". *Journal of Pragmatics* 30, 245-257.

Lenk, Uta. 1998b. *Marking discourse coherence: functions of discourse markers in spoken English*. Tübingen: Narr.

Lenk, Uta. 2005. "Discourse markers". In Östman, Jan-Ola; Verschueren, Jef (eds.). *Handbook of Pragmatics Online*. www.benjamins.com/online/hop (17 May 2006).

Levinson, Stephen C. 1983. *Pragmatics*. Cambridge: CUP.

Longacre, Robert E. 1976. "'Mystery' particles and affixes". In Mufwene, Salikoko S.; Walker, Carol A.; Steever, Sanford B. (eds.). *Papers from the 12th regional meeting. Chicago Linguistic Society. April 23-25, 1976*. Chicago: Chicago Linguistic Society, 468-475.

Müller, Simone. 2005. *Discourse markers in native and non-native English discourse*. Amsterdam: Benjamins.

Ong, Walter J. 1984. "Orality, literacy, and medieval textualization". *New literary history* 16(1), 1-12.

Östman, Jan-Ola. 1981. *You know. A discourse-functional approach*. Amsterdam: Benjamins.

Östman, Jan-Ola. 1982. "The symbiotic relationship between pragmatic particles and impromptu speech". In Enkvist, Nils Erik (ed.). *Impromptu speech: a symposium*. Åbo: Åbo Akademi, 147-177.

Redeker, Gisela. 1990. "Ideational and pragmatic markers of discourse structure". *Journal of Pragmatics* 14, 367-381.

Redeker, Gisela. 1991. "Linguistic markers of discourse structure". *Linguistics* 29, 1139-1172.

Schegloff, Emanuel A. 1982. "Discourse as an interactional achievement: some uses of 'uh huh' and other things that come between sentences". In Tannen, Deborah (ed.). *Analyzing discourse: text and talk*. Washington: Georgetown University Press, 71-93.

Schiffrin, Deborah. 1987. *Discourse markers*. Cambridge: CUP.

Schiffrin, Deborah. 2001. "Discourse markers: language, meaning, and context". In Tannen, Deborah; Schiffrin, Deborah; Hamilton, Heidi E. (eds.). *The handbook of discourse analysis*. Malden, MA: Blackwell.

Schmied, Joseph. 1998. "Discourse markers in the Lampeter Corpus of Early Modern English tracts". In Borgmeier, Raimund et al. (eds.). *Anglistentag 1997 Giessen. Proceedings*. Trier: WVT, 57-65.

Schourup, Lawrence C. 1985. *Common discourse particles in English conversation*. New York: Garland.

Stein, Dieter. 1985a. "Perspectives on historical pragmatics". *Folia Linguistica Historica* 6(2), 347-355.

Stein, Dieter. 1985b. "Discourse markers in Early Modern English". In Eaton, Roger; Fischer, Olga; Koopman, Willem; van der Leek, Frederike (eds.). *Papers from the 4th international conference on English historical linguistics. Amsterdam, 10-13 April, 1985*. Amsterdam: Benjamins, 283-302.

Stubbs, Michael. 1983. *Discourse analysis. The sociolinguistic analysis of natural language*. Oxford: Blackwell.

Svartvik, Jan. 1980. "*Well* in conversation". In Greenbaum, Sidney; Leech, Geoffrey; Svartvik, Jan (eds.). *Studies in English linguistics for Randolph Quirk*. London: Longman, 167-177.

van Dijk, Teun A. 1979. "Pragmatic connectives". *Journal of Pragmatics* 3, 447-456.

Watts, Richard J. 1989. "Taking the pitcher to the 'well': native speakers' perception of their use of discourse markers in conversation". *Journal of Pragmatics* 13, 203-237.

Widdowson, H. G. 1990. *Aspects of language teaching*. Oxford: OUP.

Widdowson, H. G. 2004. Text, context, pretext. Critical issues in discourse analysis. Malden, MA: Blackwell.

Language Acquisition revisited – a network-based approach

Lotte Sommerer, Vienna*

1. Introduction

First language acquisition has always been a fascinating area of research for linguists. As the field weaves together many diverse sources, linguists find themselves confronted with matters of biology, neurology, psychology, anthropology and cognitive science. Working with heterogeneous types of knowledge from a number of different disciplines makes it highly challenging to develop realistic and parsimonious models of language acquisition.

The discussion of language acquisition is part of the far-reaching discussion about the peculiarities of human nature. One enters the ancient battlefield of Nativism vs. Empiricism, genetic Determinism vs. the blank slate, Nature vs. Nurture, etc. Every linguistic school from structuralism and behaviorism to generativism or functionalism has positioned itself in this discussion and has developed its own assumptions about acquisition reality.

Particularly regarding syntax one central question has divided camps until today: How can children acquire syntactic rules from environmental input when it is obvious that nobody instructs babies how to put words into syntactic categories or how to structure phrases and sentences hierarchically? It is an empirical fact that children quickly reach a stage of productivity and creativity around the time of their second birthday. Then they start to produce two-word sentences with clear syntactic and semantic relations that resemble those that also characterize adult grammar. The observable 'ease' of syntactic acquisition at a stage in which the child still needs diapers and has no idea that two plus two equals four raises the question if children are excellent grammarians from the start, or if they are simply starting from scratch when analyzing their language.

Knowledge of syntax is not just memorized knowledge of a list of sentences; rather it is systematic knowledge that allows speakers to produce an infinite set of different sentences from a finite inventory of lexical

* Author's email for correspondence: lotte.sommerer@univie.ac.at

elements. Clearly, a theory of grammar must be able to explain such productivity. The generative view is that grammatical knowledge consists of a computational system of symbolic rules and that humans are genetically equipped with universal principles which enable them to learn the syntax of a certain language.

However, that view is not uncontested and one issue is particularly questionable: Even if one accepts that language consists of a set of rules and a lexicon, does this automatically imply that some of these rules have to be present at birth? Could they not also be acquired afterwards? That is to say: Is the assumption of an innate Universal Grammar really necessary? Most of the arguments for an innate foundation of grammar are not based on direct genetic or physiological evidence. Rather, the argument is that general learning mechanisms are too weak to account for the acquisition of grammar. The speech that children perceive is full of errors, false starts and slips of the tongue. Some even claim that children would get distracted and confused if they had to acquire language through the input they receive. It is argued that the child has a deficient set of data from which to work and is therefore in need of a device that gives it a head start in order to acquire the complex system behind the input.

This paper aims to show that this conclusion is unwarranted and that acquisition of syntax does not necessarily require a specialized acquisition device. It will do so by outlining a model in which at least two-word syntax is shown to be acquirable on the basis of such general capacities as imitation, categorization and analogical reasoning.

The paper is based on an MA thesis[1] and is part of a larger research project which approaches language and language change in terms of Universal Darwinism (cf. Aunger 2000; Croft 2000; Cziko 1995; Dawkins 1989; Dennett 1995; Hofstadter 1979; Ritt 2004). Among other things it is presumed that constituents of language qualify as replicators or replicating neuronal configurations implemented as patterns in the brain. These are spread through social transmission and evolve in Darwinian terms. The approach integrates findings in biology, evolution of language, connectionist studies on neural networks, Complex Adaptive System studies, Artificial Intelligence, as well as other recent findings in language acquisition studies regarding cognition, learning and theory of mind.

---

[1] *Learnability of Syntax – An Evolutionary Approach* (Sommerer, 2004) written at the English Department of the University of Vienna under the supervision of Prof. Nikolaus Ritt.

The thesis on which this paper reports wants to challenge traditional assumptions regarding the acquisition of syntax and argue for the necessity of alternative reasoning. It strongly supports non-nativist thinking i.e. it rejects the poverty of stimulus-argument. Instead it assumes that even syntactic regularities can exclusively be acquired in a learning process and get transmitted in a process of replication. A tentative network-based evolutionary model of the acquisition of simple syntax is sketched. Two-word stage data from the CHILDES (Child Language Data Exchange System) corpus is discussed in order to justify its general approach.

In this brief contribution the author aims to a) show the truly productive system that emerges in two-word stage (section 2.1) b) report on findings that reject the poverty of the stimulus-argument (section 2.2), c) link linguistic pattern recognition abilities to similar abstraction abilities observed in neural networks (section 3.1) d) put forward a network-based model (section 3.2) and e) interpret grammar as a self-adaptive system (section 4). Only a few examples from the data have been selected to demonstrate the potential of the framework.

2. Early syntactic development

The examples selected for this paper are taken from corpora out of the CHILDES archive. In the Roger Brown corpus (Brown 1973), the Sachs corpus (Sachs 1984) and the Suppes corpus (Suppes 1974) 40 files were searched for early two-word stage data. After a primary analysis of the files, nine files by four children (three girls, one boy)[2] were used for further analysis. Those nine files included examples of particular interest and the others were not as rich in examples. Additionally, sentences found elsewhere resembled the patterns in the chosen files. Not only does the collection offer a nice range of MLU's[3], but it also includes a variety of different sentence types.

---

[2] (1) Sachs, Naomi 09 – 1;10.3 (22months) MLU 1.475; (2) Sachs, Naomi 10 – 1;10.10 (22 months) MLU 1.565; (3) Sachs, Naomi 11 – 1;10.10 (22months) MLU 1.455; (4) Suppes, Nina 01 – 1;11.16 (23 months) MLU1.835;(5) Suppes, Nina 02 – 1.11.24 (23months) MLU 1.907; (6) Suppes, Nina 03 – 1;11.29 (23 months) MLU 2.274; (7) Brown, Eve 01 – 1;6 (18months) MLU 1.525; (8) Brown, Eve 02 – 1;6 (18months) MLU 1.622; (9) Brown, Adam 01 – 2;3.4 (27months) MLU 2.098.

[3] MLU = 'Mean Length of Utterance' is the measure of grammatical development. It is the average length of utterances in the samples, counted in terms of the number of morphemes.

The data illustrate the period around the second birthday, which follows earlier stages of unanalyzed chunks, 'jargon+word' combinations and holophrastic single-word utterances. During the period under investigation a truly productive system with strings of the minimum length of two emerges. At this stage, children announce the appearance, disappearance or movement of certain objects. In addition, they name owners or the properties of objects and comment on, reject or request activities that take their fancy. In the two-word stage there is a lack of syntactic and morphological markers and sentences are usually simple active declarative sentences; negative sentences and questions appear at a later stage. However, the child already has the ability to create sentences never heard before and in 95% of cases it does so without error (Pinker 1984: 123). In terms of relations expressed one can identify the following patterns:

I) Property Indication: *big drum* (Adam01 (file), 17(line), *little bug* (Nina01, 547), *red fish* (Nina01, 342);

II) Agent + Action: *Daddy go* (Adam01, 253), *man read* (Eve 02, 229), *Mary go* (Adam01, 1705);

III) Possessor + Possession: *my shadow* (Adam01, 159), *dolly shoe* (Eve01, 2034);

IV) Action + Entity: *hit ball* (Adam01, 350), *feed the rabbit* (Nina03, 46), *see truck* (Adam01, 456);

V) Recurrence, Number…: *more cookie* (Eve01, 30), *other shoe* (Naomi09, 40), *another diaper* (Naomi09, 460);

VI) Agent + Locative: *cow zoo* (Nina03, 6), *horsie zoo* (Nina03, 10), *Georgie living room* (Naomi10, 63).

VII) Action + Locative: *fall down* (Naomi09, 260), *sitting down* (Naomi09, 265);

VII) Entity + Locative: *pants off* (Naomi09, 36), *diaper on* (Naomi09, 105)

Moreover, the following interesting combinations occurred: **'want' + entity:** *want lunch* (Eve02, 3160), *want babbie* (Eve02, 3229), *want banana* (Sachs09, 149); **verb + 'it':** *go get it* (Adam01, 1572), *I did it* (Eve01 231), *taste it* (Eve01, 2244,2248,2268), *fix it* (Naomi09, 111, *like it* (Naomi09,

911); **auxiliary 'don't':** *don't push* (Naomi10, 16), *don't push recorder* (Naomi10, 35). [4]

2.1 Creativity in a truly productive system

It is important to note that two-word syntax is already truly productive, i.e. children use the words in their vocabularies in different combinations. Example (1) beautifully shows the productivity of early language.

Example 1 - Nina02/MLU 1.907[5]
665	*MOT: that's the <u>doggy's tail</u> .[6]
669	*CHI: <u>black tail .</u>
671	*CHI: <u>black tail .</u>
673	*MOT: it's a black tail ?
675	*CHI: black tail .
677	*MOT: ok # Nina # let's put on your shoes so you
	don't hurt your feet .
680	*MOT: will you step down ?
682	*CHI: <u>black doggy .</u>
684	*CHI: black doggy .

In line 665, the mother introduces the dog's tail. The child (line 669) uses the color term 'black' to refer to the tail. The mother did not use 'black' earlier on in this dialogue, so the child does not merely imitate. Thus the child is clearly creative and is able to express what she has in mind, i.e that the color of the dog's tail is black. Although the mother quickly turns to a completely new topic (line 677), the child returns to the dog and finally coins the utterance

---

[4] All the micro-sentences either correspond to Braine's list of grammatical patterns (cf. Braine 1976) or to Brown's listing of eight relational meanings in declarative constructions. Brown (1973: 189-198) proposes a list that he claimed would account for the majority of the meanings children express. These would be: I) Agent + Action, II) Action + Object, III) Agent + Object, IV) Action + Locative, V) Entity + Locative, VI) Possessor + Possession, VII) Entity + Attribute, VIII) Demonstrative + Entity. Almost all of the examples found belong to Brown's Type II, III and IV constructions. Nevertheless, I consider this categorization as an established but dubious method, as language already gets analyzed in adult terms. The patterns that occur at that stage are not necessarily based on adult categories.

[5] The normal English files have '%mor-lines' with complete part-of-speech tagging . Due to length restrictions these have been excluded here.

[6] The underlined parts in the transcription are incorporated into the text afterwards by the author and are used for emphasis.

black doggy. Apart from being an example of the child's creativity the example also suggests that the child is already able to follow the correct English word order pattern in which the adjective precedes the noun as it does not utter something like #*tail black*.

Now how does the child acquire such an ordering rule? Generally, there are three possibilities: First, categories and rules of a language may be genetic predispositions, i.e. parts of Universal Grammar. UG, also known as LAD (Language Acquisition Device), refers to the child's ability to construct or to 'invent' a grammar based on primary linguistic data, using innate predispositions as starting point (Jackendoff 2002: 70). The existence of UG is supposed to follow from the 'poverty of the stimulus-argument': environmental input a child receives is insufficient for the construction of a grammar, i.e. if something develops in the LAD that did not go in, it can only come from the structure of the mind itself (cf. Chomsky 1965, 1986b; Chomsky & Lasnik 1993; Fodor 2001; Kirby 1999; Lightfoot 1989; Radford 1990, 1999).

Especially principles and parameters theory, a widespread version of generativism, is strongly based on the assumption that children are endowed with special knowledge about permissible classes of structures or grammatical operations on those structures. The grammar of a child is supposed to consist of a set of principles valid for all languages and a set of parameters that define the range of possible differences among them. Acquiring a language means merely to set the language-specific parameters (cf. Fukui 1993; Hyams 1986; Manzini & Wexler 1988; Ouhalla 1994). In this scenario, a phrase like *black tail* or *black doggie* is acquired easily after the setting of certain innate parameters that specify possible word orderings.[7]

Secondly, semantic bootstrapping has been proposed by Pinker (1984, 1989; cf. Mcnamara 1972). Essentially it is assumed that young children have access to conceptual categories such as type of objects, relations, properties or activities on which they build word classes. In the acquisition process, these categories map directly onto the notional categories used in speech (e.g. Nouns pick out things, places or people) (Clark 2003: 201). Moreover, children know instinctively that agents are likely to be subjects, objects affected by action are likely to be direct objects and so on. Having a number of innate semantic categories, the acquisition of syntactic categories and of

---

[7] Learning as parameter setting opposes the idea of hypothesis testing. The latter view implies that the child has to sift through and decide on a range of hypotheses.

sentence structure in general, falls into place after an early rudimentary analysis (Goodluck 1991: 102).

The essential idea of semantic bootstrapping is that it offers a means of getting started on grammatical categorization as word-classes are assumed to be universal. Semantic bootstrapping is a well developed account for syntactic acquisition (cf. Bowerman 1990; Grimshaw 1981; Radford 1995) and the author believes that Pinker's ideas (compared to principles and parameters theory) come much closer to what really could be a technique to figure out a grammatical system. However, the author believes that even Pinker is not radical enough as he still assumes innate grammatical categories.[8]

Thirdly, the child may have no *a priori* access to any kind of specific information about grammar. If anything at all is innate, it is a domain general ability to recognize patterns, imitate, categorize and learn, as well as other general conceptual capacities (mnemonic and perceptual biases, voice articulation and hearing), which the child does not only use for language but for many other cognitive tasks as well. In such a framework language structure emerges from language use. (cf. Bates & Goodman 1999; Bybee 1995; Clark 2003; Croft 2001; Goldberg 1995; Givon 1995; Langacker 2000; Tomasello 2003) Clearly the last theory is most parsimonious and depends on making the fewest a-priori assumptions. It is the theory this paper wants to argue for.

2.2 Poverty of stimulus-argument rejected

The richness of 'motherese'

One view of acquisition holds that the input the child receives is often full of errors. However the empirical evidence does not really support this view as strongly as one might believe. Instead the data investigated by the author suggests that parental misunderstandings or communicative breakdown do not occur as much as expected according to the above view. In the Childes data the observable interaction between mother and child is smooth and feedback is rich. Nor is truly impoverished or ill-formed parental input to be found

---

[8] Furthermore, Pinker is a strong defender of the so-called continuity assumption, i.e "that basic linguistic representations are the same throughout all stages of child development –since they come ultimately from a single universal grammar (Pinker , 1984)."(Tomasello 2003:2). The author strongly opposes such an assumption and believes that basic linguistic representations emerge over time and develop through all stages constantly undergoing change in scope, meaning, and structure.

easily in the examples. The parental input is grammatically correct, the constructions are syntactically simple and speech is normally directed at the child in a pragmatically concrete context.

Of course, Chomsky's poverty of stimulus-argument has been attacked from different angles for many years. Attractive as generative ideas are, they need extremely strong empirical support. And this has not been forthcoming. The universal constraints Chomskyans have proposed often turned out to be assumed too hastily and without serious testing. The ideological urge to provide a back-up for whatever unifying principle has led to a selective presentation of data and far-fetched explanations (for critique cf. Croft 2001; Deacon 1997; Elman et al. 1996; Hawkins 1988; Lamb 1966; MacWhinney 1999; Tomasello 2003 etc.)[9].

Functionalist research and the social/cognitive position have presented various studies rejecting the poverty of stimulus-argument and supporting the adequacy of a strong learning mechanism within the child. Arguments in favor of a non-genetic explanation for language acquisition have focused on the following aspects:

I) the power of statistical learning mechanisms and pattern abstracting abilities (cf. Aslin, Saffran & Newport 1998, 1999; Goldberg 1995; Marcus, Vijayan, Bandi Rao & Vishton 1999; Newport & Aslin, 2000; Saffran, Aslin & Newport 1996; Saffran, Newport, Aslin, Tunick, & Barrueco 1997; Tomasello 1992, 2003)

II) the connection between general, non-linguistic development and linguistic development, (e.g. intention reading, theory of mind) (cf. Bakeman & Adamson 1984; Gopnik, Choi & Baumberger 1996; Perez-Leroux 2001; Shore, O'Connell and Bates 1984)

III) the quality of 'motherese' (cf. Clark 2003; Newport, Gleitman, & Gleitman 1977)

IV) the role of feedback

The aspect mostly discussed when it comes to the poverty of stimulus-argument is the availability of feedback. There has long been a general

---

[9] However, especially language acquisition data and various other psychological experiments have shown subversive counter-evidence. Instead of trying to take such psychological reality into account, Chomskyans, on the whole, ignored such criticisms. Instead of replying, they have tried to immunize the theory against falsification. Counter-examples were said to be peripheral and not to belong to 'core' grammar; they were declared 'contaminated' and not 'idealized' enough, or even 'dialectal'. It was also stated that counter-examples should not count unless accompanied by a fully-fledged theory (Fodor and Garrett 1966: 152). What one arrives at is a meta-grammar, programmatic rather than empirical.

consensus (Bowerman 1988; Marcus, 1993) that the child does not receive the negative feedback which seems essential for a correct system to emerge (Hoff: 2001: 245). Negative evidence has crucial implications for attempts to account for learning language from input. Mothers do correct factual errors, mispronunciations, "naughty" words, and some overregularizations, but syntactic mistakes are not corrected most of the time.

The question is however, whether a more subtle kind of correction could not exist that might help the child in its internal analysis and there seems to be evidence for that. Among other things, it was found that (a) adults are "more likely to repeat verbatim children's well-formed sentences than ill-formed ones" (Hoff 2001:245). (b) They also tend to repeat ill-formed sentences using correct alternatives. (c) Adults also ask for clarification of ill-formed sentences. (d) There is also non-verbal parental response (gestures, etc.) that helps the child to acquire the language. (cf. Bohannon & Stanowicz 1988; Chouinard & Clark 2003; Demetras, Post, & Snow 1986; Schmidt & Lawson 2002).

Aspects like frequency also influence grammatical development. There are studies which contradict the need for parental correction in general, as according to experiments, frequency and variety are the most important input properties for the child, even more important than whether the adult models a form for the child or implicitly corrects it. It is more important how often the child hears the sentence than how often it has parental support in learning it. (Valian & Casey 2003: 136). (cf Hoff-Ginsberg & Shatz 1982; Newport, Gleitman & Gleitman 1977; Richards 1994; Shatz, Hoff-Ginsberg & McIver 1989).

In short, parents do respond to grammatical and ungrammatical utterances in different ways, thus employing a more subtle correction technique which is as potentially useful to the child as overt correction. Example 2 illustrates some of this:

Example 2 - Nina 03/MLU2.274

1	*MOT: you saw animals at the zoo ?
3	*MOT: what animals did you see at the zoo ?
6	*CHI: cow zoo .
8	*MOT: what did you see at the zoo ?
10	*CHI: horsie zoo .
12	*MOT: you saw horsie at the zoo ?
14	*MOT: what else did you see at the zoo ?
17	*CHI: a lion .

19	*MOT:	you saw a lion ?
21	*CHI:	lion zoo .
23	*MOT:	you saw a lion at the zoo .
25	*MOT:	what else did you see ?
27	*CHI:	oh # baby goat .
29	*MOT:	did you see a baby goat .
31	*CHI:	poor baby goat .
33	*CHI:	a baby goat .
35	*MOT:	did you pat the baby goat ?
37	*CHI:	poor baby goat .

The mother embeds what the child has said in a more complex context like *you saw a lion at the zoo* (12, 19, 23, 29). By doing so, the child does not only hear its utterance repeated correctly which increases the frequency of positive examples, but also in a more complex way, which allows the child to figure out the position of the words used or their combinatorial possibilities. Immediately, following the child's own production, she listens to its mother's sentence, which combines *lion* and *zoo* into the pattern. Even if the child has not yet acquired the determiner or the preposition, she will be able to recognize that both times *lion* and *zoo* are preceded by unknown entities. If nothing else, this tells the child something about the combinatorial potential of *lion* and *zoo*.

- X > lion> Y> Z> zoo

3. Pattern recognition

How, then, can a child successfully learn to produce declarative sentences or NP structures, which reflect word order rules without an innate Universal Grammar. As the following example shows, pattern recognition and analogy might be sufficient.

Example 3 - Naomi 09/MLU 1.475

306	*MOT:	what's Georgie doing ?
308	*CHI:	pants on .
311	*MOT:	what's Georgie doing # honey ?
314	*CHI:	Georgie eating .

In example 3, the mother asks Naomi what Georgie is doing. The child answers the question in a meaningful way and produces a correct word order

pattern (314). The mother's input (306, 311) puts *Georgie* in first position preceding the verb twice. It is not completely implausible that even if Naomi has no knowledge about innate constraints and even if she does not understand that Georgie is an agent she will realize that 'Georgie' comes first and put the action as the second entity. In our particular example, the task for the child is even simpler, as it also copies the progressive aspect which might help, as an often used inflectional morpheme to identify 'actionness' which is in second position.

Example 4 illustrates the same principle with the acquisition of an NP.

Example 4 - Nina 01 /MLU 1.835

544	*MOT:	is that a big bug or a little bug ?
547	*CHI:	little bug .
549	*MOT:	a little bug .
551	*CHI:	big bug .
553	*MOT:	a big bug .
555	*CHI:	big mousie .

Nina produces *little bug* (547) and *big bug* (551). The adjective precedes the Noun in the NP. Now why does the child not produce #*bug big* or #*bug little*? In line 544, one can see that the mother's sentence simply does not suggest such an order. This is a creative guideline for the child. As we assume that the child gets a lot of this input, we can also see that eventually the child might be able to figure out the regularity and temporarily categorize certain items as items one could call 'Proto-Adjectives' i.e. as words coming first and others as 'Proto-Nouns' i.e. as words coming second. Afterwards, line 555 proves Nina's understanding of the generalized pattern with the creation of *big mousie*. Using creative analogy the child produces a new variant of something that she seems to have understood to represent a general pattern.

Before a network-based model will be presented in more detail, let us digress a little and have a look at connectionism and neural network modeling, as the concepts of pattern recognition and creative analogy which may underlie the example above are also central in this kind of modeling.

3.1 Neural networks and the connectionist perspective

Since the mid-1980s neural network research, also known as parallel distributed processing, has become very strong in fields such as computer science, physics, psychology, neuroscience, psychiatry, artificial intelligence, medicine and linguistics as well. (cf. Anderson 1995; Bates & Goodman

1999; Elman 1991, 1999; Elman et al. 1996; Kosslyn and König 1992; Rumelhart & McClelland 1986; Stemberger 1992; Stemberger & Handford-Bernhardt 1999).

Connectionists acknowledge that something is innate in the human brain that makes language possible, but that 'something' may not be a special-purpose, domain specific device. In their eyes language is something that we acquire with a large and complex brain that evolved to serve many different functions at once. When said that a form of behavior is innate they in fact mean that, given normal developmental experiences, it is a highly probable outcome. Grammar 'emerges'[10] in time as a result of the interaction between nature and nurture (cf. Bates & Goodman 1999).

In connectionist studies neural networks are simulated in computer programs. A neural network is a type of information processing system whose architecture is inspired by the structure of biological neural systems. Thus it is biologically motivated and has the aim of characterizing human cognitive functions computationally (Spitzer 1999: 20ff). Such networks consist of a large number of processing units which can be compared to biological neurons. Information in these networks is processed by the activation and inhibition of neurons. An artificial neural network has only three building blocks: neurodes (an artificial model of the biological neuron), interconnects (the paths or links between neurodes), and synapses (the junction where an interconnect meets a neurode). Neural networks are:

I) analog and parallel
II) an alternative to rule-based algorithms
III) learning devices
IV) capable of self-organization and generalization

ad I) Neural networks are neither digital nor serial; instead they are analog (continuous valued) and parallel, meaning that they do not contain symbolic data and algorithmic instructions in separate memory systems; instead, they store data throughout the network in the pattern of weights, interconnections and states of neurodes (Spitzer 1999: 11). Elements are not processed one after the other by a central processor, as in a PC. Rather, the complete pattern is processed at each of the neurons at the same time, which is why the brain's

---

[10] The research group has recently decided to rename their approach *emergentism* because they believe that "nature abounds with examples of emergence" (Mac Whinney 1999: preface ix).

form of processing is sometimes called parallel as well as distributed. (cf. Rumelhart & McClelland 1986).[11]

ad II) The network, in contrast to the rule-based logical serial systems contains neither rules nor calculation procedures. Its 'knowledge' resides entirely in the weights of connections among nodes, i.e. in the topology of the network as a whole.

> *[However,] although neural networks do not contain rules, what they do can be readily described by rules. This distinction may sound sophistical, but has far-reaching consequences for our understanding of ourselves.[…] Instead of rule-based algorithms working with symbols they [neural networks] consist of subsymbolic processes, which can only be described by rules and symbols to a limited degree. Moreover, the internal representations involved in these processes constantly change during these subsymbolic operations. Such rules as exist are not in the head but are post hoc ways of describing mental functions (cf. Bechtel & Abrahamson 1991, Churchland & Sejnowski 1992, Churchland 1995, Clark 1993). (Spitzer 1999: 28-29)*

ad III) Neural networks can learn. Of all the learning concepts adopted from psychology by neural network researchers, Hebbian learning is probably the best known and most widely used. Hebbian learning implies that a neuron, A, that repeatedly stimulates another neuron, B, at the times when B is firing, will have an increased effectiveness in stimulating B to fire in the future (Hebb 1949: 62).

Just as synaptic weights are not inherited and fixed in brains, they are not fixed in artificial neural networks. Instead, neural networks are trained and during this process the weights change. In other words, neural networks, like brains, learn by experience. The connection between two active neurons is strengthened through experience. Learning takes place by extracting rules from many examples, but networks learn mostly by themselves and do not need implicit instruction about rules but good examples as well as general feedback on the outputs of their states. (Spitzer 1999: 62ff).

ad IV) Neural networks are capable of generalizing and forming prototypes that self-organize. Networks also detect structures in time and therefore can make predictions. All in all, learning in neural networks demonstrates a basic principle: similarity. Similar input tends to yield similar output. If a network has learned to classify a pattern in a certain way then it will tend to classify

---

[11] When contrasted to serial processing, parallel processing has a number of significant advantages. Patterns are recognized in a single computational step that is much faster than a serial algorithm working through the patterns one after the other.

novel variations in a similar fashion. Neural networks are a kind of "analogy engine". The principle of similarity lets networks generalize their behaviors beyond the cases they have already encountered.

3.1.1 Network features

Networks have some general features which are essential, as they closely resemble phenomena that can be found in language acquisition. Three of these features are the use of 'hidden' and 'context layers' and the ability to exploit 'noise' in the system.[12] All three account for some developments in the acquisition process.

3.1.1.1 Forming categories with hidden layers

Most connectionist network models that are in use have at least three layers. An input layer to accept patterns from the outside world and an output layer to present the network's response back to the outside world. On top of that, the network includes at least one middle layer, also called a hidden layer[13].

Figur
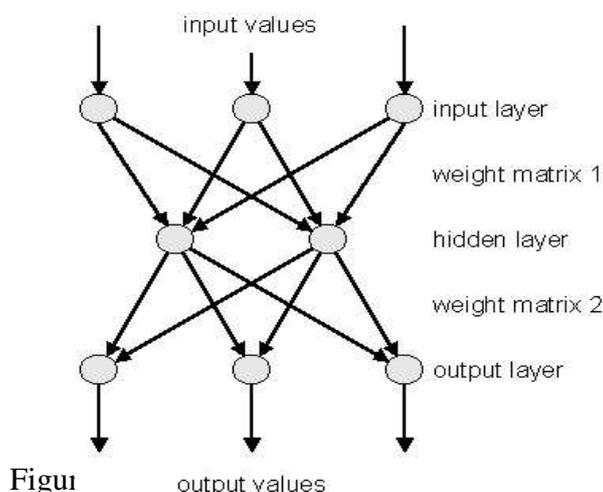
During training, the hidden layers are related to the desired outputs. This creates a map relating input to a unique output pattern. The hidden layer must

---

[12] Cf. Sommerer (2004) for other network features.

[13] As the layers between the input and the output layers are not in contact with the "external world" they are understood as something "hidden". In reality there is nothing hidden about them, which is the reason why they are called *Zwischenschichten* in German (cf. Spitzer 1999: 116)

[14] http://www.umoncton.ca/sciences/informatique/maia/in4413/projets/landry/perceptron_nm.gif

be understood as the hierarchical structure of the system. Nervous systems with such additional layers can accomplish tasks that are beyond the reach of two layer networks. It is the hidden layers that are able to generalize over the input patterns to form categories and representations of clusters of these patterns. Thus the network is made more economical and efficient.

The existence of hidden layers has been proved in biological networks and the brains of higher animals as well. The implications for language acquisition can not be underestimated as networks with hidden layers can perform abstractions and form prototypes (categories) (cf. Jones and Hoskins 1987: 156). If a network – akin to the human brain – is able to form categories, it might be capable of forming syntactic categories like Verb and Noun. As seen before, the existence of categories is considered to be essential (Pinker 1984) for a proper acquisition of syntax. Even so, there seems to be no more need to equip the child with prior knowledge about these categories if it is possible that they emerge over time.

3.1.1.2 Elman's context layers

The concept of 'starting small'

The work of Jeffery Elman[15] is of particular interest for linguists as he tries to model the acquisition of syntax taking into account that linguistic patterns unfold as temporal sequences of items (e.g. word ordering) (Elman 1991; Mozer et al. 1993). Elman networks are especially developed to represent a series of patterns in time. He added additional, so-called context layers to the hidden layers which enable the network not only to produce the correct output pattern but also to represent the temporal order of the input pattern. Neurobiologically, when we produce a sentence we use working memory (in an Elman network this is represented by the context layer) because we have to keep several words in mind in order to understand the whole sentence (cf. Just & Carpenter 1992; Carpenter et al. 1994; Petrides et al. 1993 for Neurobiological issues). As Elman networks use such 'working memory' they are also able to make predictions about the next part of speech in an incomplete sentence. This amounts to knowledge of grammar.

---

[15] Elman's network architecture has become a standard tool in subsymbolic language processing, including sentence and story understanding (Miikkulainen 1993,1996; St. John 1992, St John and McClelland 1990) and learning grammatical structure (Elman 1991; Servan-Schreiber, Cleeremans & McClelland 1991).

Of course, the crucial question is if all this can account for the acquisition of a more complex grammar as well? Essentially, simulations showed that networks were able to process complex grammar if they started simple, i.e. by learning basic sentences first, i.e. *cats chase dogs* or *mary feeds john* and only more complicated ones on top of these (cf. Elman 1993: 76). Elman points out that under some circumstances models work best when they are forced to "start small" (1993: 72) and only then undergo a developmental change which resembles the increase in working memory which also occurs over time in children. Allowing networks to "reconfigure dynamically or acquire additional nodes" (Elman 1993: 74) has been shown to facilitate learning (cf. Ash 1989; Fahlman & Lebiere 1990; Schulz & Schmidt 1991).

The important point here is that children as well as networks do not begin by mastering the adult language in all its complexity right from the start. They begin with the simpler structures, as can be seen in the given examples, and then progress incrementally until they reach adult status in their language capacity. There is even evidence that children ignore complex adult grammar not directed to them.

3.1.1.3 Noise

The existence of 'noise' is essential for networks to learn properly. Noise refers to the deformation of patterns and although the concept is derived from acoustics it can be applied to any form of signal transmission and processing. Noise always denotes an error in the signal. Reacting to noise in the system, neurons mostly produce random spontaneous activity. The greater the noise is in a network, the more likely is the chance that a neuron becomes activated that does not represent the input (Spitzer 1999: 158). Taking noise into consideration, it seems that the network might be endangered to reorganize itself in an incorrect way.

During language acquisition the child is also influenced by noise. Complex phrases it can not process in the beginning, acoustical transmission problems, mistakes (less than expected) in the mother's language, all can be taken to mean noise in the system. Here we come back to the poverty of stimulus-argument. The frequency of deficient input makes language a very noisy system, where the child gets distracted by many things and experiences difficulty distinguishing between what is worth taking into account in the acquisition process and what should not be considered.

For networks noise of this kind is not a problem. Instead they depend on it for successful learning. Actually, noise in early learning is advantageous.

Studies have shown that networks which included noise in their input learned better than those without. Elman explains:

> *The noise and instability of early learning is good news for connectionism. In contrast with previous generations of research on machine learning (where noise is always a problem), noise can be a real advantage of learning in a non-linear neural network, because it protects the system from committing itself too quickly, i.e., from falling into local minima (partial, local solutions that can prevent further and more complete learning.) (Elman et al 1996: 316).*

The challenging idea here is that an unambiguous input is not necessary for the child. It is still proficient at acquiring the correct system. Although it might face a noisy system at the beginning, it will be able to overcome the stage of confusion. The steady repetition of certain structures and items and a high frequency of good examples will enable the child to slowly figure out what qualifies as taking into account next on the basis of what it has already acquired. Former 'noise' will in time be identified as more complex structures. This reverses the poverty of the stimulus-argument completely. Deficient input is not a problem anymore but a necessary condition for successful learning. Thus confusing two-year olds is not much of a problem![16]

---

[16] It cannot be taken for granted that a network model behaves the way the real nervous system determines behavior. Still, network models can demonstrate operating principles that might be at work in natural learning. For example, the fact that network models go through similar phases resembling errors made by children suggests that children and neural networks may well learn in a similar way (Spitzer 1999: 31).

3.2 A tentative network-based model

Let us return to example 4 (p. 9) and the acquisition of the 'noun phrase' *little bug*. From an evolutionary perspective, rules represent patterns in the neural network; neural pathways that get strengthened over time. Let us assume that the entity *little* and the entity *bug* have already successfully embedded themselves in the child's brain in previous processes. This means that the neural pattern for *little* with links among conceptual nodes, phonetic nodes, and nodes for suprasegmental features has implemented itself successfully. As the child is at an immature stage, the network is not yet stably structured so that some of these pathways are not yet strongly connected, some links are still weak or tentative and some pathways still need to be added (e.g links to morphological knowledge). Separately, we can assume similar conditions for *bug*. Then the child observes the mother talk. Because of the environmental input, the child reacts and tries to repeat *little bug*. It activates *little* through links to concept nodes for size and it activates *bug* for the insect it has in mind. So there is the chance for a direct link with low stability to establish itself between *little* and *bug*. Also, the concept of *little* may be linked to the concept of *bug* reflecting the acquired knowledge that insects are typically small. If both neuronal nodes get fired one after the other (triggered through semantics) they will finally make the child express the proper sound patterns. The experienced success of such neural activity (positive feedback) strengthens the neural pathway between the two. Therefore in a first stage, the electrochemical energy will flow down the direct link. In the beginning most of the links will be arbitrary, because the neural net is only rudimentarily structured. Later on, however, both nodes will link themselves to many others.

optional pathways optional pathways

PHONE-NODES
/l/   /ɪ/   /t/   /l /

PHONE-NODES
/b/   /ʌ/   /g/

nodes for suprasegmentals

/little/   **/bug/**

nodes for suprasegmentals

CONCEPTS

[small]

CONCEPTS

[animal]

SYNTACTIC NODES

Verb

Adjective   Noun   Det

N. t. o. after

Nodes that occur
before

_____ established link .... emerging link _ _ _ _ _ destabilizing link
🌑 stabilized features   ⬯ emerging features
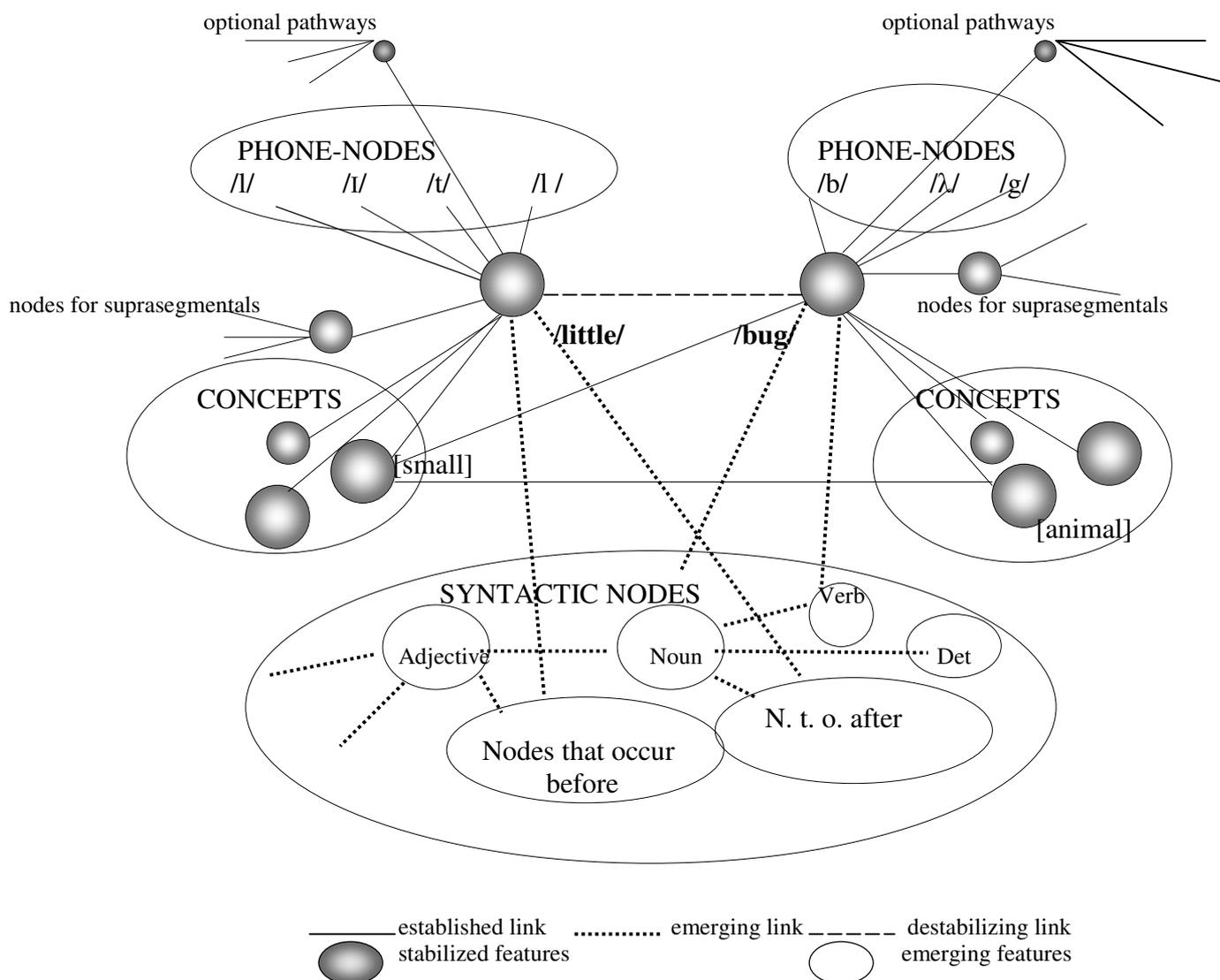
Figure 2 : Mental representation of emerging syntactic links (partially adapted from Ritt 2004: 174)

For example, the node *bug* will share its link to the concept node [animal] with many other nodes. For instance, the node [animal] will also establish a link to *mouse*. Also the child will be able to figure out that the feature [small] also fits for other animals.[17]

Obviously, this does not explain how knowledge about word order is acquired. For that three further but simple analytical steps seem to be necessary. Firstly, the child establishes many conceptual combinatorial links.

---

[17] As can be seen in the data, the child is also able to produce phrases like *little mousie* or *big mousie* (555), which is the outcome of an electrochemical connection between the *big*-node, the *little*-node and the *mouse*-node and its conceptual features [animal] and [size].

Secondly, after a while it will realize that some nodes behave syntagmatically in a similar way as others. When a child, for example, hears an utterance like *big bug* in another context, it will be able to figure out that *bug* often follows entities. This may prompt the child to categorize words into words that occur first and words that occur last. This categorization can easily be implemented in terms of links to dedicated nodes for syntagmatic sequencing, as in:

<pre>
        horsie   zoo
        Georgie   eating
        big   bug
        big   car
      \   /   \   /
       \   /   \   /
        node   node
        for words   for words
        that occur   that occur
        first   second
</pre>
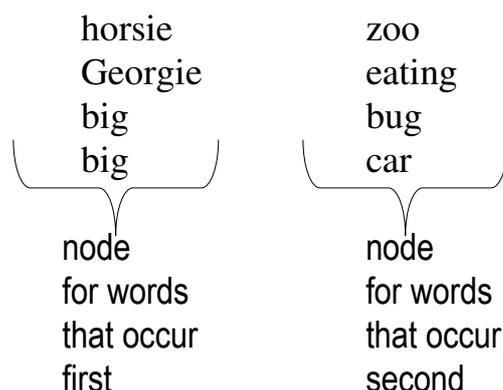
This similar pattern recognition procedure explains the examples *Georgie eating* and *horsie zoo*. Eventually, Georgie will be analyzed as 'something that occurs before' while zoo will be understood as a 'node that mostly occurs afterwards'.

In a third step, the child learns to understand that certain type of words occur in the same position and that these words are used in similar ways. The child must get to the point where it is able to see that person/things tend to occur before and that entities indicating location tend to be in final position. Then the child should be able to form categories. The formation of categories must be understood as newly emerging pathways in the neural net in which the system rearranges and simplifies itself as it will reach a state in which *bug* and other nodes become linked to a 'category specific node' that is shared by all elements that express, in utterances, as nouns do. With the emergence of such category nodes, new pathways emerge and some old pathways get destabilized or at least become relatively weaker, such as for instance, the direct link between *little* and *bug*.[18]

Another nice corollary of the perspective taken here is that it predicts children to commit errors as they begin to hypothesize about categorization and symbolic ordering as learning includes a trial and error component. As

---

[18] That such links are unlikely to disappear completely seems to be suggested by the existence of idioms, collocations and similar lexical solidarities .

the following example shows trial and error learning does indeed occur in a way that supports our perspective.

Example 5 - Eve 01/ MLU 1.525

1838	*RIC:	where is your car going ?
1841	*RIC:	is it going ?
1844	*RIC:	does it go far ?
1847	*CHI:	no [/] no .
1852	*CHI:	car coming .
1855	*CHI:	car come .
1858	*CHI:	come car .

In example 5, the mothers asks Eve during play where the toy car is going to. The child, answers with three different structures: *car coming* (1852), *car come* (1855) and interestingly the reversed form *come car* (1858). In line 1855, Eve perhaps has already understood that she has two options to refer to the concept of come: the simple and the progressive aspect. If she has no knowledge about the inflectional character of *-ing*, she may have memorized the two words as chunks and use them in free variation. What really is important is the reverse word order in 1858. There we can see that the child plays around, violates proper word order and in the end makes a 'mistake' if one wants to call it like that.

4. Syntax as a Complex Adaptive System

The way in which networks learn by reorganizing themselves represents a type of learning that has come to be attributed to so called Complex Adaptive Systems (cf. Gell-Mann 1992). CAS are supposed to acquire their structure in elaborate ways from parallel activities and the interactions of many simpler constituents including neurons in learning or genes in biological evolution (Ritt 2004: 91). A CAS has the ability to evolve, learn and to adapt to its own environment. In response to environmental feedback it is capable of altering its structure in order to become more stable. A CAS must be understood as a 'schema' which embodies information about its environment.

> *When exposed to specific environmental data – either input from sensory organs or other mental modules whose effects may amount to motivate a speaker to say something, or actual textual input – the schema which a specific competence state represents unfolds, and produces either textual output or an 'interpretation' including, and/or subsequently resulting in, new behavior. Next, the consequences of such productive or interpretative behavior feed back to the speaker's mind/brain some kind of 'evaluation' which will also affect that part of the mind/brain in which*

the speaker's competence 'resides'. If the feedback is positive, it will reinforce the original state, if it is negative it will destabilize it. Every single competence state which happens to be the basis of an unfolding will be rivaled, mind-internally by a number of other, probably rather similar states. The relative stability or strength of any particular state will correspond to the probability with which it may unfold in behavior. In such an interpretation, the 'initial state' of a linguistic competence has probably to be conceived of a set of rivaling schemata with similar degrees of (relatively low) stability. Which of them actually get expressed will be more or less a matter of chance, resulting in what may appear from the outside as relatively unsophisticated and purposeless behavior. From the very beginning, however, the rivaling schemata will come under 'selection pressure' from the feedback incurred by individual unfolding (both 'productive' and interpretative'). In response to such pressures, the population of rivaling schemata will assume a more complex organization, until the system reaches a comparably stable state. (Ritt 2004: 99)

In case of individual learning the set of schemata can be neuronally implemented patterns of thought that represent ways of interpreting the world. The results of the behavior which is generated by those patterns in particular situations can influence how those patterns fare in competition with others (Gell-Mann 1992: 11).

If we consider a child's competence as a complex adaptive system which evolves and learns, the question about language acquisition can be addressed in the following way:
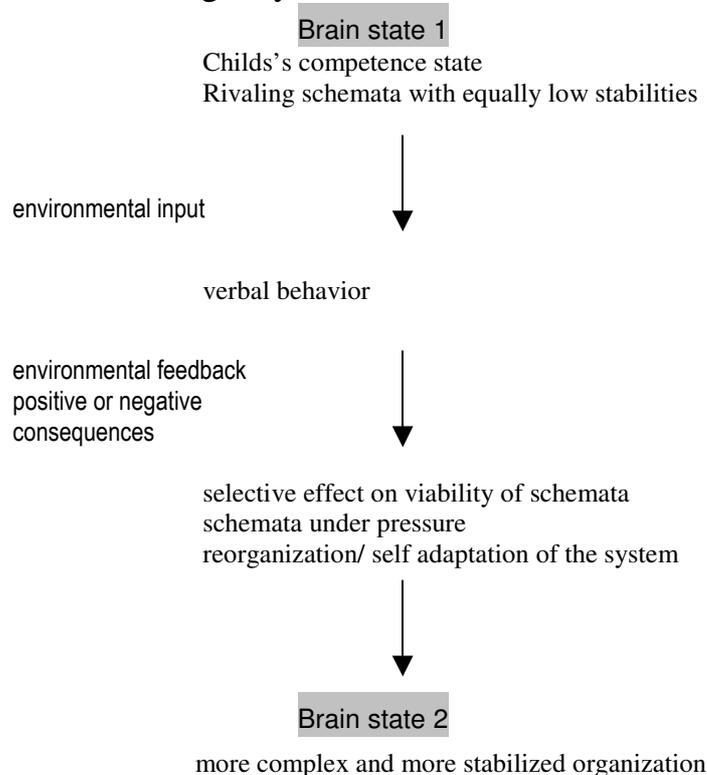

Figure 3: Complex Adaptive System (adapted from Gell-Mann 1992: 11)

The child's brain state reaches a complex and fairly stable organization because its verbal behavior triggers environmental feedback that has a selective effect on the rivaling schemata a child comes up with during the acquisition process. The system undergoes a slow process of adaptive selforganization.

4. 1 Rivalling schemata

4.1.1 Unanalyzed chunks

A case which suggests how 'selfadaptation' and 'selection among schemata' might take place within the mind of a child is the transitional stage in which unanalyzed lexical chunks get reinterpreted by the child. Thus, for example, the phrase *See you later* is used by Naomi flawlessly at a very early stage. If we take the child's immature grammatical state into consideration, we cannot assume that Nina has already acquired the syntax to construct the phrase. Rather *seeyoulater* is one of the unanalyzed chunks that the child attaches to a certain concept and memorizes as a complete word. Although *seeyoulater* consists of more morphemes, it probably gets memorized as one soundsequence with one concept (use it when you leave /alternative to *bye*). In evolutionary terms, the sequence implements itself as one neuronal chunk in the brain. At this competence state, *seeyoulater* is the outcome of only one neuronal complex expressing itself.

Similarly, in the following example *gogetit* can be assumed to be implemented in the child's mind as an unanalyzed chunk with a meaning of something like "grasp object + movement".

> Example 6 - Adam01/ MLU 2.098
> 1569   *MOT: isn't that part of your trailer ?
> 1572   *CHI: go get .
> 1574   *MOT: <u>you go get it</u> .
> 1576   *CHI: <u>Mommy go get it</u> .
> 1578   *CHI: <u>go get it</u> .

In 6, Adam and his mother are playing with a ball that rolls away. In line 1576, Adam coins the phrase *Mommy go get it*. He immediately repeats the chunk again (1578). As his mother uses *go get it* in line 1574 herself, we may

hypothesize that Adam repeats it as an unanalyzed chunk as he has not produced the structure on its own but imitated it.[19]

But what happens later on in acquisition? At a more advanced stage the syntactic knowledge of the child will have increased enormously. In the case of *gogetit*, Adam will come to the point when he will have figured out that *go*, *get* and *it* are entities of their own, which occur in various constructions and not only in connection with each other. The child will also figure out that *get* can be preceded by other nodes, e.g. *You get it*. *Mama gets …*, and that *it* replaces objects. *Get the ball*, *get your shoes*,…the child will also realize that sentences like #*the girl go get it* are ill formed as he reaches a point in which he successfully has acquired the 3rd person singular marker +*s*.

We therefore face two rivaling schemata. Firstly, the earlier unanalyzed chunk and second the new emerging system with many interconnections representing advanced knowledge about structures, regularities and grammatical features. In time, the separate *go*, the *it* and the *get* will have been triggered much more often than the *gogetit*-chunk. It is possible that the latter will destabilize itself, as we must assume that the construction *go get it* will finally be produced through the triggering of the three separate entities. Their neural pathways will have strengthened and widened through steady electrochemical flow during the expression of various phrases like *I go*, *I get the ball*,…Especially as *gogetit* can not be integrated in a more advanced grammatical system (#*He gogetit the ball*).[20]

4.1.2 Misinterpreted input

Is there ambiguous input in parental speech and can this lead to misinterpretation? Without doubt there are situations in which input indeed can lead to wrong conclusions about the grammatical system to be learned.

As has been pointed out, a network while analyzing data, sometimes tends to fall into so-called local minima and ends up at partial solutions. Generally, the system produces an output that mirrors previous input. If such input,

---

[19] His mother also uses the construction in Adam02 /4004, 4010, 4042.

[20] In the case of *seeyoulater*, things are not as clear. There is a strong argument in support of the idea that *seeyoulater* as a memorized chunk – akin to one meme – does not disintegrate. Firstly, *seeyoulater* is used quite frequently in speech and it can always be used successfully as an alternative to *bye-bye*. It is an idiomatic good bye phrase and as its use does not encounter negative feedback due to grammatical incorrectness, it is possible that next to the individual clusters *see*, *you* and *later*, *seeyoulater* survives as a single neurnal cluster in the child's brain.

especially in child language, is only partially understood by the child, it can be assumed that sometimes this leads to incorrect conclusions. However, systems are able to reorganize themselves and overcome such temporary, intermediate stages.

One situation, in which such a misinterpretation might take place is the case of the pronoun *it* in object position. All children in all the files analyzed, repeatedly uttered verb phrases like: *find it, fix it, get it, close it, taste it, like it, did it,…*Often, verbs like *find, fix, get, close, taste* are followed by *it*. This pattern is not unexpected as the valence of these verbs implies the necessity of an object position. Since *it* is a pronoun that can stand for many open class nouns, the pattern V+*it* will obviously be much more frequent than any individual V+N combination.

In the following example, Eve's mother wants her to find her father's ashtray and expresses that twice in line 4320 and 4323 with the phrase *you find it* and again in line 4329 with the question *did you find it?* repeating the daughter's utterance *find it* (4326). Here, it can be observed that the mother does not ask something like *Did you find the ashtray?*, but substitutes the noun with the pronoun *it*.

Example 7 - Eve 01/ MLU 1.525

4314	*MOT: look on the floor .
4317	*CHI: find it .
4320	*MOT: you find it .
4323	*MOT: you find it .
4326	*CHI: find it .
4329	*MOT: did you find it ?
4332	*MOT: bring it here and I'll fix it for you .
4335	*CHI: fix it .

Afterwards, in line 4332, the mother uses the phrase *fix it*. The child repeats it immediately. Also in the Suppes corpus, Nina's mother regularly uses *it* with verbs like *fix, close*.

Example 8 - Nina 02/ 1.907

1192	*MOT: what happened to the train ?
1194	*MOT: did it fall over ?
1196	*CHI: fix it .
1198	*MOT: do you want me to fix it for you ?
1200	*CHI: fix it .
1202	*MOT: here .

The data suggests the following: Firstly, the mother's input is full of *it*-object constructions. Secondly, the child successfully copies them and produces *it* in its patterns. The fact that the mother hardly uses these verbs in alternative constructions and mostly in combination with *it* and the fact that the children never produce patterns that embed these verbs into other combinations, in which they, for example, precede a noun or another entity, suggest that the child might not understand these phrases as consisting of two entities but rather as unanalyzed patterns like *getit, fixit, closeit* or *tasteit*. This, in a second step, might even lead the child to the false interpretation that an *–it* ending indicates something like 'verbhood'.

This suggests that *tasteit*, *fixit* and *closeit* might implement themselves as single chunks in the child's brain. The emerging syntactic system thus reaches a temporary competence state with some stability supported through the input given by the parent. If such a scenario really takes place, one should be able to find ill-formed patterns like *#I taste it banana* in the transcripts. And indeed, examples can be found in the data that mirror such interpretations. Adam, in example 9, and Naomi, in example 10, both put an additional object behind *it*, which indicates that they are unable to understand that *it* already fills the necessary object slot.

> Example 9 - Adam01/ MLU 2.098
> *CHI: Mommy get it ladder .
> *CHI: Mommy get it my ladder .
>
> Example 10 - Naomi09/ MLU 1.475
> 905   *CHI: jacket .
> 908   *CHI: need it jacket .

These two examples are not the only instances where such sentences are formed by the child. Naomi (file 14/ line 678) and Eve (file 04/ line 444) still form the same sentences later on in their development[21].

How does the child get over such a stage and end up at a correct grammar? In the process of acquisition, the child will receive more and more input. On the one hand, the mother will use the verb in different and more complex contexts which allow a different interpretation; on the other hand, the child will receive some kind of feedback for sentences like the one above. Such a feedback will have positive or negative consequences as it has a selective effect on the viability of the current state. Thus, the schemata will get under

---

[21] 678   *CHI: I want it hug. / 444   *CHI: see it paper.

pressure and the system will reorganize itself. Finally, the child's neural network of interconnected memes, reaches a more complex and more stabilized organization.

The given examples have pointed out that language acquisition is not a straightforward process in which one parameter after the other gets set. Errors and misinterpretations do occur and the input sometimes indeed leads the child astray. The system needs time and enough correct input to set the right interconnections. If this does not take place from the beginning, the network is fortunately capable of readaptation – another reason why the system does not have to be up and running from the start in order to guarantee the outcome of syntactic acquisition that is observable in all children: a full working adult grammar.

5. Conclusion

This paper has sketched a network-based model of simple two-word stage syntax and tried to explain some observable child language patterns with the knowledge from other disciplines. The main idea put forward in this paper is that language development is the result of the "interaction between general learning mechanisms and a richly structured environment" (Plunkett & Schafer 1999: 51). The model thus directly counters both the claimed inadequacy of the learning mechanism and the poverty of stimulus assertion that are the foundations of Chomskyan nativism. In short, it opposes representational innateness, especially principles and parameter theory. General cognitive learning mechanisms are found to be strong enough to achieve syntactic aspects of grammatical knowledge. Secondly, the claim is rejected that language experience simply does not provide sufficient feedback. Thirdly, it is argued that the syntactic system "emerges" as a result of the interaction between nature and nurture and thus works like a complex adaptive system (CAS). Such systems are capable of evolving in a Darwinian way. They exploit accidental variations among their constituents and go on to select, in response to environmental pressures, those which allow the systems to become adapted and thus more stable. Fourthly, it is shown that rules are not really rules as such but are regularities in a neural network which the child is able to filter out, select and abstract from the input because it has an extraordinary ability to do so, based on creative analogy and a clever learning strategy (the child has to sift through and decide on a range of hypotheses). Syntax is present in the weights of the connections in a neural network and this network (thus the rules) constantly undergoes change.

The author is well aware of the fact that so far network modeling of learning has historically been extremely simple and largely inadequate to explain complex behavior systems. Nevertheless, researchers have underestimated the information that might be computed from linguistic input and also the remarkable learning and computational capacities of infants. Recent empirical findings must change the balance of attention towards a network-based perspective. After all, there is a widespread rejection of explanations that produce outcomes indirectly and without stipulation and explain X by saying X is innate (Aslin, Saffran & Newport 1999: 362).

References

Anderson, J.A. (1995). *An introduction to neural networks*. Cambridge, MA: MIT Press.

Ash, T. (1989). Dynamic node creation in backpropagation networks. *Connection Science* 1, 365-375.

Aslin, R.N., Saffran, J., & Newport, E. (1998). Computation of conditional probability statistics by 8-month-old infants. *American Psychological Society* 9 (4), 321-324.

Aslin, R.N., Saffran J.R., & Newport, E.L. (1999). Statistical learning in linguistic and non-linguistic domains. In B. MacWhinney (ed.). *The emergence of language* (359-380). New Jersey: Lawrence Erlbaum.

Aunger, R. (2000). *Darwinizing culture: the status of memetics as a science.* Oxford: Oxford University Press.

Bakeman, R. & Adamson, L. (1984). Coordinating attention to people and objects in mother-infant and peer-infant interactions. *Child Development* 55, 1278-89.

Bates, E., & Goodman, J.C. (1999). On the emergence of grammar from the lexicon. In B. MacWhinney (ed.). *The emergence of language* (29-80). New Jersey: Lawrence Erlbaum.

Bechtel W., & Abrahamson A. (1991). *Connectionism and the mind*. Cambridge, MA: Blackwell.

Bohannon J.N., & Stanowicz L. (1988). The issue of negative evidence: adult responses to children's language errors. *Developmental Psychology* 24, 684-689.

Bowerman, M. (1988). The "no negative evidence" problem. How do children avoid constructing an overly general grammar? In V.A. Hawkins (ed.), *Explaining language universals*. Oxford: Blackwell.

Bowerman M. (1990). Mapping thematic roles onto syntactic functions. Are children helped by innate linking rules? *Linguistics* 28, 1253-1289.

Braine, M.D. (1976). Children's first word combinations. *Monographs of the Society for Research in Child Development*, *41, 1, Serial NO. 164*.

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

Bybee, J. (1995). Regular morphology and the lexicon. *Language and Cognitive Processes* 10, 425-455.

Carpenter, P. A., Miyake, A., & Just, M. A. (1994). Working memory constraints in comprehension: Evidence from individual differences, aphasia, and aging. In M.

Gernsbacher (ed.), *Handbook of Psycholinguistics* (1075-1122). San Diego, CA: Academic Press.

Caudill, M., & Butler, C. (2000). *Naturally intelligent systems*. Cambridge: MIT Press.

Chomsky, N. (1965). *Aspects of a theory of syntax*. Cambridge, MA: MIT Press.

Chomsky, N. (1986b). *Barriers*. Cambridge, MA: MIT Press.

Chomsky, N., & Lasnik, H. (1993). The theory of principles and parameters. *Syntax* 9.1, 506-570.

Chouinard, M., & Clark. E. (2003). Adult reformulations of child errors as negative evidence. *Journal of Child Language* 30, 637-669.

Churchland P.S. & Sejnowski, T.J. (1992). *The computational brain*. Cambridge, MA/ London: MIT Press.

Churchland, P. M. (1995). *The engine of reason, the seat of the soul: a philosophical journey to the brain*. Cambridge: Cambridge University Press.

Clark, E.V. (1993). *The lexicon in acquisition*. Cambridge, England: Cambridge University Press.

Clark, E.V. (2003). *First language acquisition*. Cambridge: Cambridge University Press.

Croft, W. (2000). *Explaining language change. An evolutionary approach*. Harlow: Longman Linguistics Library.

Croft, W. (2001). *Radical construction grammar. Syntactic theory in typological perspective*. Oxford: Oxford University Press.

Cziko, G. (1995). *Without miracles: universal selection theory and the second Darwinian revolution*, Cambridge, Mass: MIT Press.

Dawkins, R. (1989). *The selfish gene.*(2nd ed)*. Oxford: Oxford University Press.

Deacon, T.W. (1997).*The symbolic species. The co-evolution of language and the brain*. New York: Norton.

Dennett, D.C. (1995). *Darwin's dangerous idea. Evolution and the meanings of life*. New York: Touchstone.

Demetras, M.J., Post, K.N., & Snow, C.E. (1986). Feedback to first language learners. The role of repetitions and clarification questions. *Journal of Child Language* 13, 275-292.

Elman, J. (1991). Distributed representation, simple recurrent networks, and grammatical structure. *Machine Learning* 7, 195-225.

Elman, J. (1993). Learning and developmenti n neural networks: the importance of starting small. *Cognition, 48*, 71-99.

Elman, J. (1999). The emergence of language: A conspiracy theory. In. B. MacWhinney (ed.). *The Emergence of language* (1-28). New Jersey: Lawrence Erlbaum.

Elman, J., Bates, E., Johnson, M., Karmiloff-Smith, A., Parisis, D., & Plunkett, K. (1996). *Rethinking innateness: A connectionist perspective on development*. Cambridge, MA: MIT Press.

Fahlman, S.E., & Lebiere, C. (1990). The cascade-correlation learning architecture. In D.S. Touretzky (ed.). *Advances in neural information-processing systems II* (524). San Mateo: Morgan Kaufman.

Fodor, J. (2001). Setting syntactic parameters. In M. Baltin & C. Collins (eds.). *The handbook of contemporary syntactic theory* (768-824). Malden, MA.

Fodor, J. & Garrett, M. (1966). Some reflections on competence and performance. In J. Lyons & R. Wales, (eds.). *Psycholinguistic papers*. Edinburgh: Edinburgh University Press.

Fukui, N. (1993). Parameters and optionality. *Linguistic Inquiry, 24*, 399-420.

Gell-Mann, M. (1992). *Complexity and complex adaptive systems*. In J.A. Hawkins & M. Gell-Mann. (1992). *The evolution of human languages* (3-18). New York: Addison-Wesley.

Goldberg, A.E. (1995). *Constructions: a construction grammar approach to argument structure*. Chicago: University of Chicago Press.

Gopnik, A., Choi, S., & Baumgartner, T. (1996). Cross-linguistic difference in early semantic and cognitive development. *Cognitive Development* 11, 197-227.

Goodluck, H. (1991). *Language acquisition. A linguistic introduction*. Oxford: Blackwell.

Grimshaw, J. (1981). Form, function, and the language acquisition device. In C.L. Baker & J.J. McCartey (eds.). *The logical problem of language acquisition*. Cambridge, MA: MIT Press.

Hawkins, J. A. (1988). Explaining language universals. In J.A. Hawkins (ed.), *Explaining language universals* (3-28). Oxford: Blackwell.

Hebb, D.O. (1949). *The organization of behavior*. New York: Wiley.

Hoff, E. (2001). *Language development* . (2nd ed.). Stanford: Wadsworth.

Hoff-Ginsberg, E. (1986). Function and structure in maternal speech. Their relation to the child's development of syntax. *Developmental Psychology* 22, 155-163.

Hoff- Ginsberg, E., & Schatz, M. (1982). Linguistic input and the child's acquisition of language. *Psychological Bulletin* 92, 3-26.

Hofstädter, D.R. (1979). *Gödel, Escher, Bach: an eternal golden braid*. New York: Basic Books.

Hyams, N. (1986). *Language acquisition and the theory of parameters*. Dordrecht: Reidel.

Jackendoff, R. (2002). *Foundations of language. Brain, meaning, grammar*. Oxford: Oxford University Press.

Jones, W. P., & Hoskins, J. C. (1987). Back propagation: A generalized delta learning rule. *Byte Magazine*. October.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: individual differences in working memory. *Psychological Review* 99, 122-149.

Kirby, S. (1999). *Function, selection and innateness. The emergence of language universals*. Oxford: Oxford University Press.

Kosslyn, S.M., & Koenig, O. (1992). *Wet Mind: The New Cognitive Neuroscience*. New York: Free Press.

Lamb, S. (1966). *Outline of stratificational grammar*. (Revised Edition). Washington. D.C: Georgetown University Press.

Langacker, R. (2000). A dynamic usage-based model. In M. Barlow & S. Kemmerer (eds.), *Usage-based models of language*. Stanford: SLI Publications.

Lightfoot, D. (1989). The child's trigger experience: degree-0 learnability. *Behavioral and Brain Sciences* 12, 221-275.

MacWhinney, B. (1999). The emergence of language from embodiment. In B. MacWhinney (ed.), *The Emergence of language* (213-256). New Jersey: Lawrence Erlbaum.

MacWhinney, B. (1999). *The Emergence of language*. New Jersey: Lawrence Erlbaum.

Manzini, M.R., & Wexler, K. (1988). Parameters, binding theory and learning ability. In *Linguistic Inquiry* 18, 413-444.

Marcus, G.F. (1993). Negative evidence in language acquisition. *Cognition* 46, 53-85.

Marcus, G.F., Vijayan, S., Bandi Rao, S., & Vishton, P.M. (1999). Rule learning by seven-month-old infants. *Science* 283, 77-80.

McNamara, J. (1972). Cognitive basis of language learning in infants. *Psychological Review,* 17, 1-13.

Miikkulainen, R. (1993). *Subsymbolic natural language processing: An integrated model of scripts, lexicon, and memory*. Cambridge, MA: MIT Press.

Miikkulainen, R. (1996). Subsymbolic case-role analysis of sentences with embedded clauses. *Cognitive Sciences* 20, 40-73.

Mozer, M. C., Smolensky, P., Touretzky, D., Elman, J., & Weigend, A. (eds.). (1993). *Proceedings of the Connectionist Models Summer School 1993*. Lawrence Erlbaum Publishers.

Newport, E.L. & Aslin, R.N. (2000). Innately constrained learning: Blending old and new approaches to language acquisition. In C. Howell, S.A. Fish & T. Keith-Lucas (eds.). *Proceedings of the 24th Annual Boston university conference on language development (1-21)*. Somerville, MA: Cascadilla Press.

Newport, E.L., Gleitman, H., & Gleitman, L. (1977). Mother, I'd rather do it myself. Some effects and noneffects of maternal speech style. In C.E. Snow & C.A. Ferguson (eds.). *Talking to children: language input and acquisition* (109-150). Cambridge, England: Cambridge University Press.

Ouhalla, J. (1994*). Introducing transformational grammar. From rules to principles and parameters*. London.

Pérez-Leroux, A.T. (2001). Subjunctive mood in Spanish child relatives at the interface of linguistic and cognitive development. In K. Nelson, A. Aksu-Koç & C. Johnson (eds.). *Children's Language: Volume 11*, (69-93). Hillsdale: Lawrence Erlbaum Publishers.

Petrides, M., Alivisatos, B., Meyer, E., Evans, A.C. (1993). *Functional activation of the human frontal cortex during the performance of verbal working memory tasks*. Proc Natl Acad Sci USA 90:878-882.

Pinker, S. (1984). *Language learnability and language development*. Cambridge, MA: Harvard University Press.

Pinker, S. (1989). *Learnability and cognition: the acquisition of argument structure.* Cambridge, MA: MIT Press.

Plunkett, K., & Schafer, G. (1999). Early speech perception and word learning. In M.Barett (ed.), *The development of language* (51-72). East Sussex, England: Psychology Press.

Radford, A. (1995). Phrase structure and functional categories. In P. Fletcher & B. MacWhinney (eds.). *The handbook of child language* (483-507). Oxford: Blackwell.

Radford, A. (1990). Syntactic theory and the acquisition of English syntax: the nature of early child grammars of English. Oxford: Blackwell.

Radford, A. (1999). *Syntactic theory and the structure of English: a minimalist approach*. Cambridge: CUP.

Richards, B.J. (1994). Child directed speech and influences on language acquisition: methodology and interpretation. In C. Gallaway & B.J. Richards (eds.). *Input and interaction in language acquisition* (74-106). Cambridge, England: Cambridge University Press.

Ritt, N. (2004). *Selfish sounds*. Oxford: Oxford University Press.

Rumelhart, D.L., & McClelland J.L. (1986). *Parallel distributed processing: explorations in the microstructure of cognition*. Cambridge, MA: MIT Press.

Sachs, J. (1983). Talking about there and then: the emergence of displaced reference in parent-child discourse. In K.E. Nelson (ed.). *Children's language, Vol.4*. (1-28). Hillsdale, NJ: Erlbaum.

Saffran, J.R., Aslin, R.N., & Newport, E.L. (1996). Statistical learning by 8-month-old infants. *Science* 274, 1926-1928.

Saffran, J.R., Aslin, R.N., Newport, E.L, Tunick, R.A., & Barrueco, S. (1997). Incidental language learning: Listening (and learning) out of the corner of your ear. *Psychological Science* 8, 101-105.

Schulz, T.R., & Schmidt, W.C. (1991). A cascade-correlation model of balance scale phenomenon. *Proceedings of the Thirteenth Annual Conference of the Cognitive Science Society*. Hillsdale, NJ: Erlbaum.

Schmidt, C., & Lawson K. (2002). Caregiver attention-focusing and children's attention-sharing behaviour as predictors of lateral verbal IQ in very low birth weight children. *Journal of Child Language* 29, 3-22.

Servan-Schreiber, D., Cleermans, A., & McClelland, J.L. (1991). Graded state machines. The representation of temporal contingencies in simple recurrent networks. *Machine Learning,* 7, 161-194.

Shatz, M., Hoff-Ginsberg, E., & McIver, D. (1989). Induction and the acquisition of English auxiliaries. The effects of differentially enriched input. *Journal of Child Language* 16, 121-140.

Shore, C., O'Connell, B., & Bates, E. (1984). First sentences in language and symbolic play. *Developmental Psychology* 20, 872-880.

Snow, C. E. (1999). Social perspectives on the emergence of language. In B. MacWhinney (ed.). *The Emergence of Language* (257-276). New Jersey: Lawrence Erlbaum.

Spitzer, M. (1999). *The mind within the net: Models of learning, thinking, and acting*. Bradford Books.

Sommerer, L.(2004). *Learnability of syntax – an evolutionary approach*. Unpublished MA thesis, University of Vienna.

Stemberger, J. P. (1992). A connectionist view of child phonology. Phonological processing without phonological processes. In C.A. Ferguson, L. Menn & C. Stoel Gammon (eds.). *Phonological Development* (16-189). Timonium, MD: York Press.

Stemberger, J.P., & Handford-Bernhardt, B. (1999). The emergence of faithfulness. In B. MacWhinney (ed.), *The Emergence of Language (*417-446). New Jersey: Lawrence Erlbaum.

St. John, M.F. (1992). The story gestalt: A model of knowledge-intensive processes in text comprehension. *Cognitive Sciences* 16, 271-306.

St. John, M.F., & McCelland, J.L. (1990). Learning and applying contextual constraints in sentence comprehension. *Artificial Intelligence* 46, 217-258.

Suppes, P. (1974). The semantics of children's language. *American Psychologist*, 29, 103-114.

Tomasello, M. (1992). *First verbs: A case study of early grammatical development*. Cambridge, England: Cambridge University Press.

Tomasello, M. (2003). *Constructing a language. A usage-based theory of language Acquisistion*. Cambridge Mass: Cambridge University Press.

Valian, V., & Casey, L. (2003). Young children's acquisition of wh-questions: the role of structure input. *Journal of Child Language* 30, 117-143.

Electronic sources

CHILDES Database:
http://childes.psy.cmu.edu/.
(http://childes.psy.cmu.edu/manuals/0intro.pdf

Implementing Change: Applying new assessment procedures at the Department of English and American Studies, Vienna University

Susanne Sweeney-Novak*

1. Introduction and Background

The UniStG 1997 required Austrian universities to rewrite or redesign their existing curricula until 2002. This request initiated a concerted effort by colleagues at the English Department of the University of Vienna to review the then existing language programme;[1] to investigate students' strengths and deficits on entering university to study English[2]; to identify the specific needs of students wishing to study a subject which is conveyed through the medium of English.

The first step in devising a new programme was to identify the **aims**[3] to which it should be geared, i.e. the competencies successful learners in higher education should have developed in the course of their studies. These competencies or abilities are listed in detail in the new curriculum but can be summarised as follows in their relevance to the language programme, and in particular to the test under review:

- ability to understand, critically analyse and evaluate complex situations within the field of research
- ability to apply strategies and techniques for problem solving

---

* Author's email for correspondence: susanne.sweeney-novak@univie.ac.at

[1] A special tribute must be paid here to Pat Häusler-Greenfield, who was deeply involved in developing the present language programme. Not only had she worked on other language and Cultural Studies syllabuses, but her expertise on language teaching and learning, especially in EAP, ensured an informed approach to syllabus design and development with an outcome appropriate to the needs of students and departmental staff alike.

[2] Particular mention must be made of Heinz Ribisch' work on "can do" and "can't do" statements of first-semester students, which was an important source for drawing up a list of "can do" targets for the first year of the language programme. Dr Ribisch is Lektor at the English Department.

[3] The terms "aims" and "objectives" (below) are taken from Widdowson (1983: 7).

- ability to think on a multi-dimensional level
- ability to integrate and conceptualise new ideas

In order to meet these long-term aims, the working group first identified deficiencies of first-semester students in the following areas with a view to remedying them by devising an appropriate language programme:

- functional competence
- vocabulary
- study skills and learning strategies
- language and text awareness (weak to non-existent reading/writing skills; poor listening skills)
- discourse competence (levels of formality, style, register or appropriateness)
- specific speaking techniques such as presentation, discussion [4]

Therefore, a new language programme was needed, one which was "more effective, more appropriate to the students' current and future needs and more principled, consistent and transparent." Although the language programme covers three years, it is with the first year and its assessment procedure that this paper is concerned. Before these specific issues are addressed, the following table will briefly illustrate the overall development within the language programme:

1st year	Integrated Language and Study Skills (ILSS 1 & 2)	Skills based	Developing skills and strategies to cope effectively in an academic context (B2/C1 of the Common Eurpean Framework)
2nd year	Language in Use (LIU 1 & 2)	Skills based AND Discourse based (emphasis on supra-sentential features, rhetorical organisation, patterning and context)	Understanding characteristic features of texts; Analysing what makes for effective communication; Producing material in a (limited) range of genres, showing awareness of the lexical, syntactic and stylistic choices involved (C1)
3rd year	Advanced Integrated Language and Study Skills (AILS 1 & 2)	Skills based AND Text based	Translation and advanced text analysis; Ability to work with and in English in a manner appropriate to the medium chosen, audience and purpose. (Effective Operational Proficiency / Mastery; C2)

To combat the above-mentioned deficiencies, the first level of the new language programme (ILSS 1 & 2), in contrast to previous syllabuses, is a **skills-based** programme. An important influence in defining skills and

---

[4] This information is taken from an internal document on the development of the first-year language programme called "ILSS 1& 2 Background".

devising the language programme was Jordan (1997)[5]. Jordan defines skills for academic purposes not only in terms of the four **language skills** in their macro and micro manifestations, but includes **cognitive skills** (language and learning skills and strategies), **study skills**, and **communication skills** (effectively conveying information).

Given the framework of a skills-based approach, a detailed syllabus was developed to form the common basis for teaching in ILSS 1 and 2. These papers are referred to as the **grids**. They describe in detail the course content, the skills and sub-skills and strategies necessary for students to "make the transition from intermediate level of proficiency … to a more reflective, flexible and targeted advanced level appropriate to a course which uses English as the medium to convey complex and sophisticated ideas" (*ILSS 1 & 2 Background*).

To fulfil the aims of this programme, and on the express wish of the Department, a standardize test for all students at the end of Year 1 was developed. In this respect, the test functions as the **agent of change** without which transparency of assessment and comparability of results would not be guaranteed. This test, whose principles, development, format and standardization procedures will be explained below, is called the **Common Final Test**, referred to in the course of this paper as **CFT**.

This paper will, furthermore, outline the general principles of test design and stages of test development as described in the testing literature; it will then link these theoretical concepts to the CFT and explain the steps taken to ensure the CFT's validity and reliability. Comprehensive statistical data will be supplied to explain further the standardization procedure applied in the test development process.

2. Test purpose

The CFT was first implemented in the summer semester '03 for students who had begun their English studies at the Department in the winter semester '02/03 and who were the first to follow the new curriculum. The CFT is taken by students at the end of their second semester in the Integrated Language and Study Skills (ILSS) component of their studies.

As this article is concerned with testing in general and one test in particular, the term "test" should be briefly defined here. In the testing literature and in common usage other terms are used to indicate judgements

---

[5] Other influences on the programme development were: Swales (1990) and Tribble (1996).

being made: evaluation, assessment and tests[6]. Often these are used interchangeably; rarely definitions are given. For brevity and clarity's sake, these three terms are defined as follows:

EVALUATION
refers to judgement of a whole programme e.g. the effect of a new coursebook

ASSESSMENT
is the part of evaluation which focuses on the learner;
it refers to various processes through
which outcomes are assessed/judged
formative[7]

TESTS
are the tools used in the
evaluation/assessment process
summative

These terms are arranged in a hierarchical fashion indicating the breadth of their application. It shows that tests are limited in their scope and restricted by time. These limitations do not apply to other assessment procedures, notably the portfolio, which is characterised by work over time.

Continuous assessment is applied in ILSS 1 & 2. The CFT, however, constitutes 50 per cent of the final grade of ILSS 2; students must receive a pass in order to receive a positive grade for that semester.

The purpose of the CFT is to ensure that the learning of relevant academic skills as specified by the syllabus and manifested in the grids has taken place. It is, as such, regarded as an **achievement** test.[8] As the term "achievement test" implies, it looks back at previous teaching to see whether **course objectives** have been met. However, in its wider implications it contains

---

[6] The term 'exam' or 'examination' is used synonymously with 'test', "with no apparent clear distinction in meaning between the two" (Davies 2000).

[7] The terms formative and summative are also closely linked with the purpose of assessment and tests. 'Formative' would stress aiding development, giving feedback, monitoring learning processes, whereas 'summative' stresses judgement on product.

[8] The terms achievement test, proficiency test or placement test are used to indicate which purposes tests serve. Textbooks on testing all use this terminology. For a more detailed description on test purpose and use cf. Bachman (1990: 54-61).

strong elements of a **proficiency test** in that it elicits skills and abilities which meet the needs of students in their future academic careers, thus adding a predictive dimension to the test results. (Common European Framework 2001: 183)

3. Principles of test design

Every assessment procedure should follow the three principles of test design. These are
- Validity
- Reliability
- Practicality

Validity in the testing literature is seen as a complex concept but an inevitable prerequisite for a good test. Most importantly, validity is closely linked to the purpose for which a test is designed. We have established above that the CFT is an achievement test; it can, therefore, be called valid if it contains major elements of the syllabus; in other word, if it tests the desired objectives of the syllabus. It would be less valid if, to be successful, a student had to have specialist knowledge in certain areas which are not specified in the syllabus objectives.

In section 4 below, certain types of validity are mentioned of which Construct Validity is the most complex one. It involves a definition of the abilities, language or otherwise, relevant for teaching and testing, and it involves an understanding of what the theories behind these abilities are. Content validity stresses the link between test and syllabus; as we can see from the previous paragraph, construct and content validity in achievement tests are often rolled into one as the theoretical understanding of abilities has determined the content of the syllabus.

Reliability is concerned with measurement, scoring and rating. Very often reliability conflicts with validity. A writing test, for example, can be highly valid in that it tests directly what it claims to test, but falls short of reliability if different raters were to mark or grade its outcome differently. On the other hand, a test of grammar could be highly reliable but not valid if it claims to test language competence. (See below the points made about TOEFL and IELTS).

Practicality is also important in test design as one has to be aware of the limitations set by institutions, by time and by financial constraints to achieve what might be considered the "perfect test". (See below arguments for not including listening and speaking in the CFT).

4. General stages of test development[9]

The following framework illustrates, in general, the stages any test development and design should follow, such as answering questions as to what it is one wants to test how this should be tested (validity) and what the results tell us about the test and the candidates (reliability).

1st	Conceptualisation Which behaviour should the test elicit? What should the test consist of? Reviewing literature	Validity Construct validity: to which extent is the test based on sound theoretical principles, and to which extent does the test accord with the aims defined in the curriculum and the objectives of the syllabus
2nd	Selecting tasks which elicit behaviour identified in Stage 1 Drawing up Test Specification: test structure, timing, length, task types, weighting, assessment criteria (scoring method)	Content validity: to which extent is the sampling of texts, tasks and items representative of the construct
3rd	Piloting Pre-testing – feedback from colleagues Trialling – feedback from colleagues and students	Concurrent validity: how far does a pilot test warrant similar results as an existing test
	Analysis of test items	Reliability The extent to which a test measures consistently

The following chapters explain the development of the CFT in adherence to the above framework.

5. Establishing Validity

5.1. Looking for a test format: integrative vs. discrete point

When developing the assessment system for ILSS 1 & 2, it was felt that in order to ensure a "principled, consistent and transparent" system, a common test was required which would be taken by all students at the end of their second semester. Given the fact that developing a standardized test takes about two years, and given the fact that such a development is expensive, it was clear that neither time nor money were available to develop a departmental test. Once having established which abilities and skills were relevant for academic life (i.e. defining the construct), two international tests

---

[9] I am indebted for this framework to Alan Davies' lecture series on Testing held at the English Department, Vienna University in the summer semester '97.

were considered as possible models and analysed as to which best tested these skills and abilities.

The two international English tests relevant for academic life are TOEFL (mainly for American universities) and IELTS (mainly for British and Australian universities). In comparing these two international tests, it became obvious that the content of the IELTS test suited the language programme's requirements much better than did the TOEFL test. One of the reasons was that the TOEFL construct does not conform fully to the concept of language ability as envisaged by the new curriculum. For example, TOEFL is still based to a large extent on indirect testing (a classic example of this could be the testing of pronunciation through a multiple choice pencil-and-paper test). It still uses discrete items (i.e. items which can be answered without understanding of context or for which no context is necessary and which form the basis of the structural approach to language description). Furthermore, TOEFL started only in the nineties to include actual writing tasks in its test battery following demands from universities who argued that students who had done well on the TOEFL test could not write.

The history of TOEFL vs. the Cambridge tests is the history of objective/psychometric, i.e. highly reliable testing, vs. subjective and highly valid testing as Bernard Spolsky describes in his history of the development of objective language testing (Spolsky 1995) and in his introduction to the comparability study of TOEFL and Cambridge tests:

> *That a psychometrically pure test is reliable is unquestionable. What remains to be shown convincingly is the nature and relevance of what it measures. The quest for perfection in tests (like the quest for the holy scale in the assessment of more integrative language skills) founders ultimately, I believe, on the multidimensional, interactive, dynamic, affective and contextualised nature of language proficiency. (Bachman et al 1995: 13)*

The designers of the new syllabus view language proficiency as **integrative**[10]: integrating language knowledge (the criterion which discrete point testing would tap such as discrete knowledge of grammar or vocabulary) and the ability to use this knowledge for purposeful communication and performance. This means that they do not see language proficiency as a product broken down into discrete component parts for the sake of being easily measurable on the basis of which judgement can be made

---

[10] The term "integrative" is used here more generally and as opposed to discrete item testing; it is not used in the sense of J.W.Oller , who in the late seventies strongly advocated the use of cloze tests and dictation as being the most integrative types of test.

as to the effective use of language for a defined purpose. It is for this reason that the CFT has no specific grammar, vocabulary or language-in-use section, but evaluates a candidate's use of language knowledge from within the reading and the writing part of the test. This approach, it is believed, conforms to what some scholars in their criticism of existing models of communicative competence say on the issue of the multi-divisibility of language proficiency:

> *If communicative tests are to move forward they will need to address the problem of feature isolation ..., whereby features of language use are analysed out and performance necessarily distorted because performance is not a collection of features but an integrated interplay between them. (McNamara 2000:20f)*

Widdowson (2003: 168ff), likewise, criticizes the most influential model of communicative competence at present developed by Bachman (1990) and Bachman & Palmer (1996) for presenting a framework which isolates features of competence but fails to show the inter-dependency of those features. He says of models of communicative competence[11]:

> *The assumption behind their [the models'] development from Hymes to Bachman and Palmer seems to be that the more differentiations the analysis can yield, the greater its operational value. But this surely presupposes the very discrete item view of language testing that these models are designed to discredit. (Widdowson 2003: 170)*

From the theoretical standpoint of an integrated view of language performance, the adoption of the IELTS format suitable for a common final test seemed the best option available.

The development of the IELTS test is a very impressive feat[12]. What is important is that the needs for academic purposes were defined by the IELTS test developers (see the description of the needs analysis in Clapham (1996)), from which an instrument for testing the abilities to meet these needs was developed. As there was no time to develop a departmental assessment tool – the new syllabus was implemented in the WS 02/03 and the first CFT was first held in SS03 - a decision had to be made to adapt the IELTS test for the purpose of a common departmental test. This was a good choice from a practical point of view as well, as there is a wealth of preparatory material on the market which could be used in ILSS classes to ensure familiarity with test tasks and format.

---

[11] An exhaustive overview of models of communicate competence (except for the Bachmann & Palmer model of 1996) can be found in Weir (1990).

[12] For a brief summary see Alderson et al (1995: 23), for an extensive report see Bachman et al (1995) and Clapham (1996) and for a historical comparison of TOEFL and IELTS see Spolsky (1995).

5.2. The CFT format

IELTS covers all four language skills: listening, speaking, reading and writing. The CFT, however, only tests reading and writing. This has been criticized, but the reason is a practical one (third principle of test design). It was simply not feasible to develop a listening test for the CFT as it would have implied recording spoken texts, editing them and using a recording studio, all of which was not possible for financial reasons. The CFT does not have a speaking part either, because it was felt that speaking would be covered in the students' career at a later stage anyway, and that testing the spoken language proficiency of 200 to 300 students per semester would have been excessively time-consuming.

Deficits in beginning students' reading and writing skills had been diagnosed in the preparatory work on the syllabus. It was generally felt that reading and writing were important skills for university students and devising a test that would contain one receptive and one productive skill might even out the restriction posed by practicality.

The CFT consists of a Reading and a Writing part for which 120 minutes are allocated. The **Reading** part of the CFT consists of three reading passages of about 800 words in length each; 60 minutes are reserved for task completion. The texts are taken from textbooks, magazines, journals and the internet. The principles for selecting an article for the test are:
- the topic should be of general interest for the students;
- the topic should be of relevance to arts students;
- the topic should be in some way related to the course of study;
- the proposition or propositions of a text should be clearly structured and organised on paragraph and text level (Urquhart & Weir 1998: 141ff).

So far the test developers have not had to take special care as regards cultural bias when selecting a text, because the testing population is seen, more or less, as a homogenous group. But cultural bias will certainly be an issue for future text selection if the number of foreign/international students studying at the Department increases.

The **Reading** part of the CFT tests, for example, students' ability to
- understand texts globally (surveying for gist) and in detail
- look quickly for specific information
- make inferences
- distinguish between main and supporting ideas
- summarise by extracting salient points

- understand and critically evaluate the writer's point of view .

All these skills and strategies are important for someone who has to read and understand a great amount of text in an academic context. In addition to these skills, the CFT also tests whether a candidate understands the meaning and function of specific advanced vocabulary (though not in the form of the classic multiple-choice task format) and rhetorical devices with which a writer conveys meaning (e.g. modality).

Writing Task 1 is a transformation task: students are presented with graphic information which they have to describe. For task completion 20 minutes are recommended.

This task was chosen by IELTS developers as a text type because the abilities needed to fulfil this task are regarded as academically relevant:

- cognitive ability to understand non-verbal information
- cognitive ability to interpret, compare and contrast non-verbal information and evaluate its relevance
- language competence to report findings clearly, precisely and economically.

These abilities are seen as the basis for successful scientific research and reporting necessary in an academic context; the skills and strategies acquired in class to fulfil Writing Task 1 successfully are those skills which can be transferred to academic situations in which, for example, detached analysis is undertaken.

Writing Task 1 is the most controversial task in the CFT. Students do not like describing graphic information; teachers do not enjoy teaching it. Despite repeated explanation as to the reason for the inclusion of this task, rejection is strong. This is indeed a serious problem as it affects the **face validity** (the acceptability of a test as an appropriate instrument of assessment) of the CFT. Nevertheless, this task can serve as a perfect example for what is meant by ability in the above sense and task purpose in teaching and testing.

In his discussion of the terms "aims" and "objectives" of ESP and GPE course design, Widdowson (1983: 6-7) explains:

> *By objectives I mean the pedagogic intentions of a particular course of study to be achieved within the period of that course and in principle measurable by some assessment device at the end of the course. By aims I mean the purpose to which learning will be put* after *the end of the course. Thus a course may have as one of its objectives the development of the ability to carry out certain specific experiments in chemistry, but the aim of this exercise would refer to a more general capacity for problem solving and rational enquiry which learners could apply to later experience even if they had no further contact with chemistry for the rest of their lives.*

This quotation illustrates beautifully the thinking behind the inclusion of graphic representation in the CFT. If we believe that "clear" thinking (or in Widdowson's terms capacity for problem solving and rational enquiry) is essential for tertiary education (aim), and not only reserved for the natural sciences, and that it is, furthermore, an ability worthy of development, then we must find ways of training it (objectives). In testing terms we would equate aims with the construct of a test and objectives with its content.

For **Writing Task 2** students have to write an impromptu text, incorporating the features of an argumentative or expository essay. Students are presented with a proposition/a problem and have to fulfil a specific discursive task on it (discuss, argue, describe advantage and disadvantage). The prompt will look like this:

Time

Register

You should spend about 40 minutes on this task.

Present a written argument or case to an educated reader with no special knowledge of the following topic.

General Topic, Topic sentence & support

Nowadays there is no realistic way of cutting yourself off from the communications blitz. We are all exposed to more communication than anyone would have believed possible fifty years ago. In some respects this has clearly improved the quality of our lives, but at the same time, modern methods of communication have threatened many things we once believed were important.

Proposition Problem

Function Task

Say whether you see instant communication as a blessing or a curse.

You should write between 250 and 300 words. You should support your arguments with examples and relevant ideas.

Your text will be assessed according to content, organisation, range of vocabulary and accuracy.

Assessment Criteria

Compared to Writing Task 1, Writing Task 2 tests different abilities and skills. The test-taker's competencies are closely linked to those required for the reading tasks, albeit **productive** in nature. Writing Task 2 focuses on the student's ability to develop and support an argument and in so doing employ acquired language competencies relevant for the task such as presenting cause and effect, problem and solution, comparison and contrast.

5.3. Authenticity of text and task

The points made above about the CFT format also indicated which abilities are being tested in the Reading and Writing parts of the test. These abilities, as mentioned above, have been defined as relevant for academic study. Yet a sceptic might easily wonder how performance on a test can be linked to real-life performance. Can anyone, in good faith, believe the predictions made on the basis of test results? The task of designing a test is to simulate real-life situations which are or will be of relevance for the target group. Although students will have to read considerably longer texts than the three 800-word texts they are faced with in the CFT and likewise will have to write considerably longer texts than the 250-word composition, the abilities and skills employed by candidates in the test situation are representative samples of future behaviour in the real world, in this case in the academic context.

Authenticity, for pedagogic purposes, is generally understood as the employment of texts featuring real-life properties such as format and textual organisation (recipe vs. newspaper article vs. letter to the editor) and pragmatic principles (communicative function, register and discourse). On this basis such texts can be called realistic. Similarly, tasks in a testing situation, for example reading tasks, can and should reflect the potential real-life purposes of the readers and test the skills and abilities required in the relevant context. McNamara (2000: 8f) points out, however, that although "materials and tasks in a test might be realistic … they can never be *real*." The only authenticity or reality for that matter is the testing situation itself.

Texts for the three reading passages in the CFT are chosen for their propositional and heuristic qualities, their lexico-grammatical complexity and their general purpose to engage the reader. The text types chosen are, however, of a limited range: semi-academic texts chosen from popular specialist magazines, reports of research findings taken from newspapers and the like. The restrictions given by the test specifications regarding the length of a text often mean that these genuine texts have to be altered. This could mean shortening them, or removing passages not necessary for the overall propositional frame. Sometimes very low-frequency words are replaced by high-frequency ones, and on rare occasions sentences or definitions are inserted to facilitate, for example, cultural understanding.

Whereas the authenticity of a reading test can be established quite easily, the authenticity of a writing task in an academic context is problematic. Weigle (2002: 52) points out the relative inauthenticity of the timed

impromptu essay by suggesting that real-life features of writing in an academic context are missing:

- use of source material as input (e.g. assigned reading)
- mostly not timed or speeded
- awareness of audience (the students' instructor)

While it is true that these limitations do affect writing performance, it is also true that authenticity can be found, in Writing Task 2 for example, in the way students respond to the prompt: positioning oneself, constructing an argument, challenging the prompt. The third limitation quoted above is offset by the fact that students are aware of the assessment criteria which are binding for all markers (for details see section 7.3).

6. Test development

6.1. Test Development: Reading

Each stage in the development of the CFT is based on team effort. This co-operation can take the form of looking for appropriate texts, arguing for or against the inclusion of certain texts, giving feedback on the efficacy of test items, and proof reading. Team effort is important in test development (Bachman & Palmer 1996: 165), because a single test developer can easily develop tunnel vision and not view her work objectively enough. However, although most colleagues are willing to assist, they can do so only to a limited extent taking the unpaid nature of this work into consideration.

The following diagram illustrates the development of the Reading Test from the initial stage of searching for appropriate texts to the final version ready to be put into the item bank.

```
┌─────────────────────────────────────────────────────────────┐
│ Selection of texts according to test specification          │
└─────────────────────────────────────────────────────────────┘
                              │
                              ▼
┌─────────────────────────────────────────────────────────────┐
│ Item writing and test compilation                           │
└─────────────────────────────────────────────────────────────┘
                              │
                              ▼
┌─────────────────────────────────────────────────────────────┐
│ Pre-testing with native speakers and non-native colleagues and feedback │
│ Pilot testing in ILSS 2 classes (Practice Test);  feedback from students │
└─────────────────────────────────────────────────────────────┘
                              │
                              ▼
┌─────────────────────────────────────────────────────────────┐
│                  Statistical Item Analysis:                 │
│ Facility Value, Discrimination Index, Reliability Index and feedback results │
│ Poor items are discarded and some items are re-written      │
└─────────────────────────────────────────────────────────────┘
                              │
                              ▼
┌─────────────────────────────────────────────────────────────┐
│ Revised Test                                                │
│ Lay Out,  Proof Reading,                                    │
│ Put into bank for use in later semesters                    │
└─────────────────────────────────────────────────────────────┘
```

6.2. Test development: Writing

For **Writing Task 1** a graphic is chosen which should be of relevance to the students. It is, in most cases, taken from newspapers or current publication. So far, some areas of difficulty for students have been identified, such as confusing titles or redundant information, and wherever possible, these will be avoided.

The general topic for **Writing Task 2** has in most cases been of relevance to current affairs (for example, "manned space flight" after the destruction of a space shuttle in 2004, or "the nanny state" after the introduction of smoke-free zones) or of relevance to the students' own lives (introduction of study fees; internal or external assessment). It is clear that not every topic may be of interest to all students and that some students may find it very hard to come up with arguments in connection with specific topics, but a modicum of

awareness of the world around us should help students fulfil Writing Task 2 satisfactorily as far as content is concerned. We are in this respect luckier than test developers for international tests in that we have a fair idea of the background knowledge of our students. However, as mentioned above, supposing the number of international students in the Department increases, a greater awareness of cultural-specific background will be required.

7. Establishing Reliability

7.1. Assessment: Reading and Writing

At present, every item on the Reading Test carries one point. No item is weighted and 60 percent of correct answers constitute a pass (see below the present discussion on the efficacy of the cut-off point). The Reading part constitutes **50 per cent** of the overall grade on the test.

For Writing Task 1 and 2 a detailed set of criteria (see below for an example of the marking scale for Organisation) for evaluation has been developed. These analytical scales should ensure fairness and transparency in grading, but they also provide insight into the writing construct and the writing abilities required of the students. For example, they show, as said repeatedly in this paper, that linguistic competence is seen as an integral part of writing ability. Testing it discretely and then drawing conclusions about a candidate's writing ability would not be in accordance with the writing construct of the CFT.

The Writing Tasks are weighted. Task 1 carries **20 per cent** and Task 2 carries **30 per cent** of the overall mark on the test.

7.2. Reliability – Reading

The analysis and statistical validation of the practice tests (trial runs – compare stages of test development above) give an indication as to whether a test will accord in standard to previous tests.

To ensure reliable results for Reading, the statistical analysis of trial and actual tests is of utmost importance as no test can be valid if it is not reliable. The following tables give an indication of the analysis procedures every trial test and every actual test is subjected to and what information is to be gathered from these analyses.

Statistical Test Analysis

For any test, wherever set or taken, a simple analysis of the data would help to get a clear picture of test takers and of the efficacy of the test itself. Rarely is this done in classroom situations. However, in developing a test which is supposed to measure consistently, statistical analysis is of vital importance in order to make sense of data, add objectivity, highlight trends and relationships in the data and make meaningful predictions.

Statistical formulae are omitted here. The uninitiated but interested reader is referred to Hughes (1989), who caters for those teachers who would normally shy away from anything to do with numbers and figures, to Alderson et al. (1995), whose more complicated approach provides invaluable insight into the intricacies of test analysis, and to Brown (1988), who does not focus on language testing only, but on statistics and research design in general.

In the following, the main steps of analysis will be described and the results to date will be reported. These steps include

A: Item analysis: Facility value and discrimation
B: Central tendency and dispersion
C: Standard error of measurement and reliability coefficient

A. Item Analysis: Facility Value and Discrimination

Facility Value

The first step in test validation is to see how "good" individual items are. An item is considered appropriate when it shows the right level of difficulty, i.e. when it is neither too easy nor too difficult for the target group, and when it discriminates between good and weak candidates.

The level of difficulty of an item is shown by the facility value (FV). This represents the percentage of students answering an item correctly. The higher the FV, the easier an item is. In developing a test one would ideally aim at producing items with a FV of 0.5, with more correct answers in the top group (i.e. those 50 per cent of candidates who have higher overall scores on the reading test than the other 50 per cent i.e. the bottom group). But as this is unlikely to be achieved in every case, FVs of a range between 0.2 and 0.8 are acceptable for us. More precisely, the experience we have gained from the test runs so far tells us that we should not have a higher mean FV on the whole test than 0.7 and ideally would stay between 0.65 and 0.7.

One of the problems we encounter is the fact that items sometimes perform differently in trial runs to the actual test. This could be due to the greater number of candidates between trial and actual test, or it could be due to the increase of stress on the candidates in the actual test, but we have yet to learn what other variable(s) might account for the different item analysis results between trial and actual test. Future correlation studies might give a better insight into this phenomenon.

Table 1 Facility Value (on the difficulty of sub-tests and overall tests)

	Reading 1	Reading 2	Reading 3	**Reading Test Mean**
SS03	0.71	0.75	0.64	**0.70**
WS03/04	0.70	0.87	0.77	**0.78**
SS04	0.80	0.80	0.70	**0.76**
WS04/05	0.63	0.66	0.59	**0.62**
SS05	0.76	0.65	0.76	**0.72**
WS 05/06	0.69	0.65	0.66	**0.67**

As Table 1 shows, the Reading Part of the CFT of the winter semester 04/05 was the most difficult reading test so far, but as far as level of difficulty is concerned, it is more in keeping with our target.

In recent semesters we have started to compare and count the number of easy items (**E**, facility value .65+), items of middle difficulty (**M**, facility value .40 – .64) and difficult items (**D**, facility value below .40). These figures add additional information on test difficulty and give us a better understanding of the impact and precision of the cut-off point.

Table 2

	WS04/05	**SS05**	**WS05/06**
E (65+)	24 (25 cut-off point)	32 (25 cut-off point)	24 (27 cut-off point)
M (40-64)	11	8	15
D (below 40)	6	1	4
No. of items	41	41	43

This form of item analysis for the winter semester '04/05 and the summer semester 05 was done retrospectively and showed that the great number of easy items in the summer semester's test for which the cut-off point of 25 was too low. A more conscious approach was adopted in the winter semester 05/06 when the cut-off point was raised by one point above 60 per cent. Clearly, in future this information must also be taken into account to ensure that the cut-off point is fairly set.

Discrimination

In analysing the efficacy of an item, it is also important to check whether an item discriminates well between good and weak candidates. We check among other things for items with a negative discrimination index. These are items which were answered correctly by a greater number of weak students (the bottom half) than good ones. After the trial run, items with a negative discrimination index are eliminated. We also check whether items with a high facility value (easy items) might, nevertheless, be good discriminators or whether items with a low facility value (difficult items) discriminate well, in which case we include these items in the final version of the CFT.

Certain items frequently are of a very low facility value. These are very often items which test the understanding of modality, of low-frequency vocabulary, and of the writer's attitude. These items, however, are very important in testing language on an advanced academic level. We will therefore include them if their discrimination index tells us that only the very good students were able to answer them.

B. Central Tendency and Dispersion

While facility value and discrimination above evaluate individual items on a test and can indicate whether a test is of an appropriate level of difficulty for the target group, the following statistical data of central tendency and dispersion provide information on group behaviour. They do, in addition, offer useful information for the standardization of a test.

Central Tendency

Traditionally, teachers will register the average grade of a class or the average point a class will achieve in a test. This single figure can, however, be very misleading. For example, two groups can have the same average, but are totally different in their distribution of grades. For example, the average or mean of 2.5 is reached by two classes with 5 students each: Class 1: 1 1 1 3 5 and Class 2: 2 2 2 2 3. But as we can see, the distribution of grades varies. Class 2 is much more homogeneous, Class 1 has a wider range of grades with a greater number of excellent student at the top. It is, therefore, important to record additional figures of central tendency:

- **the mean** indicates the average (number of correct items divided by the number of students)
- **the mode** is the figure reached by the greatest number of students (Class 1=1; Class 2=2)
- **the median** is the middle figure between the top and bottom half of a group (above Class 1=1; Class 2=2).

Ideally these figures will cluster together.

Table 3: Comparison the figures of central tendency: the mean, the mode and the median.

Central Tendency	SS03 41 items	WS03/04 43 items	SS04 43 items	WS04/05 41 items	SS05 41 items	WS05/06 43 items
Mean	**29.4**	**33.53**	**33.62**	**26**	**29.93**	**28.6**
Mode	**31**	**40**	**34**	**27.5**	**30/32 bimodal**	**34**
Median	**30**	**34**	**33**	**27**	**30**	**29**

Table 3 is in accordance with the above data on the facility value of the reading tests. It confirms that the reading test in the winter semester '04/05 was difficult. It is interesting to see, however, that in all semesters the mode and the median are above the mean, which indicates, as one would expect in an achievement test, a cluster in the area of greater ability.

Dispersion

Table 4 compares the figures of dispersion: the **range**, which indicates how widely the scores are spread out; and **standard deviation (SD)**, which indicates the average amount by which each student's score deviates from the mean.

Table 4

Dispersion	SS03	WS03/04	SS04	WS04/05	SS05	WS05/06
Range[13]	**20-38 = 18**	**13-43 = 30**	**16-42 = 26**	**10-35 = 25**	**16-41 = 25**	**14-42 = 28**
SD	**4**	**5.7**	**4.7**	**5**	**4.8**	**7**

A wide range as in the winter semesters '03/04 or '05/06 is commendable, as this shows that good students were able to reach the top end of the range and that weak students were further away from the cut-off point. The latter was not met in the summer semester 03, where the lowest point of 20 is too near the cut-off point of 24, nor was the former met in the winter semester '04/05 as none of the students were able to reach the top score.

---

[13] The range was calculated according to Alderson (1995: 93), i.e. highest score minus lowest score, while Brown (1988: 66f) uses the formula highest score minus lowest score plus one.

C. Standard Error of Measurement (SEM) and Reliability Coefficient (RC)

The Standard Error of Measurement (SEM) helps to make predictions about a student's **true score**[14]. For example, an SEM of 3 tells us that we can be 68% certain that a student who had scored 28 points on the test would, when taking the test again, be in the range of 25 to 31. (Two standard errors of measurement would even give a 95% certainty of the true score). The higher the reliability of a test, the lower the SEM will be.

To know the SEM is important for borderline cases. The SEM may also be an argument for keeping to the decision taken for the CFT pass/fail requirement as it was in the winter semester '03/04. This means that a negative Reading Part should not automatically lead to an overall negative mark on the CFT (although one could argue that a very low reading score should fail a student).

The Reliability Coefficient (RC) is an important statistical figure in standardizing test results. It will tell test designer to what degree a test measures consistently and how meaningful the results are. A low RC is the results of errors within the test. Ideally, we should aim at an RC of 0.7 or above.

Table 5 compares the reliability coefficient of the CFTs to date. The low RC of the summer semester 03 was the result of a limited trial period. The subsequent RCs are fairly satisfactory and even better than in the trial tests.[15]

Table 5 Standard Error of Measurement (SEM) and Reliability Coefficient (RC)

	SS03	WS03/04	SS04	WS04/05	SS05	WS05/06
SEM	5	3 (2.85)	3 (2.7)	3 (3.05)	3 (2.8)	3 (3.08)
RC	0.49	0.75	0.66	0.64 Trial 0.54	0.66 Trial 0.62	0.81 Trial 0.65

---

[14] The **raw score** is the number of items the candidate answered correctly and which, in turn, is converted into a grade; the true score, however, indicates within which range the candidate might fall if he/she were to take the same test again.

[15] The formula for estimating reliability is Kuder-Richardson 21 (Alderson et al 1995: 282f)

7.3. Reliability – Writing

Unfortunately we can provide only very limited statistical data on the reliability of the CFT's writing tasks as we lack the software necessary to analyse data in detail. Nevertheless, we can, to some extent, comment on the steps taken to ensure the best possible reliability in the assessment of the writing tasks. The assessment of writing tasks is, contrary to assessing reading tasks, to a great extent subjective and dependent on the assessor.

In order to ensure reliability in the assessment of the two writing tasks in the CFT, two important steps have been taken. First, the compilation of a fairly detailed, **analytic marking scale** for both writing tasks, and second, extensive **rater training** after the inception of the CFT.

In the developmental stage of the CFT, ILSS course teachers discussed thoroughly the advantages and disadvantages of **holistic vs. analytic scoring**. Most teachers will be used to **impressionistic scoring** of learners' written texts, which means awarding a score without any reference to explicit criteria. The drawback of such an approach is clear: there is little diagnostic feedback to the learner as to which areas of competence need to be improved. Holistic scoring, on the other hand, works with a set of criteria, but these are broad, overall descriptors of levels of proficiency. Again there is little feedback to learners. Weigle says in her very useful summary of various scoring methods about holistic scoring:

> *This is especially problematic for second-language writers, since different aspects of writing ability develop at different rates for different writers: some writers have excellent writing skills in terms of content and organization but may have much lower grammatical control, while others may have an excellent grasp of sentence structure but may not know how to organize their writing in a logical was (2002: 14).*

Given the disadvantage of traditional impressionistic scoring, it was felt that in order to ensure the target of transparency and comparability of results an analytic scoring frame should be developed.

In Writing Task 1 (graph description) the following aspects are assessed:

Task Fulfilment	5 points
Organisation	4 points
Vocabulary and Sentence Structure	5 points
Length	1 point

Each aspect in turn is described in a 5-point scale (or 4 points in the case of Organisation). The **'Task Fulfilment'** scale, for instance, looks like this:

5 p	The student seems to have fully grasped the contents and all important points relevant to the task are mentioned. There is no redundant information: the student is able to condense/paraphrase if necessary.
4 p	The one or the other point of information may have been misunderstood, but all important points are mentioned. Some redundant information is included, and there is little or no condensation/paraphrasing.
3 p	Some points misunderstood and/or misrepresented, but most important points of information are mentioned; redundant information is included (= linear, non-selective description), but the reader who had not seen the graphic would still get the correct general idea of the graph.
2 p	Some points misunderstood and/or misrepresented so that the reader who had not seen the graphic would get a wrong impression. Important points of information are omitted. Moreover, information is presented in a linear, non-selective fashion or in a confusing way.
1 p	Key points of content are misunderstood and/or misrepresented (e.g. figures, trends) so that the reader who had not seen the graphic would get a totally false impression. Moreover, important points of information are omitted and information is presented in a linear, non-selective fashion or in a confusing way.
0 p	Task and contents misunderstood.

In Writing Task 2 other aspects are assessed:

Content, Argument & Evidence	5 points
Organisation: Coherence & Cohesion	5 points
Vocabulary (range & depth)	5 points
Grammatical patterns	5 points
Punctuation	1 point
Spelling	2 points
Length	2 points

The scale for **'Organisation'**, for example, looks like this:

5 p	The text is organised into paragraphs with clear topic sentences and the text can be seen to be structured into introduction, main body and conclusion. A variety of linking devices is used between paragraphs. Logical relations between sentences and overall method of development (e.g. result, cause & effect, comparison and so on) are indicated by the appropriate use of lexis, conjunctions and discourse markers.
4 p	The text is organised into paragraphs and the topic of a paragraph is obvious to the reader. The text can be seen to be structured into introduction, main body and conclusion. A limited range of linking and cohesive devices is used.
3 p	The text seems to be organised into paragraphs but there is no internal organisation within these. OR: The text is not organised into paragraphs although there is a discernable introduction, main body and conclusion. As a result of the lack of paragraphing the textual development is difficult to follow. There is some use of cohesive devices.
2 p	The text is not organised into paragraphs and there is little evidence of structure / framing: unsuccessful introduction, conclusion etc. There is hardly any use of cohesive devices beyond "and/but".
1 p	The text is not organised into paragraphs and there is no introduction or conclusion. There is little or no use of cohesive devices beyond "and/but".
0 p	There is no apparent structure.

The marking scales as they stand at present have been subjected to continuous revision on the basis of their efficacy of application. They are still "work in progress" as we find that some vague terms like "some or little" are open to interpretation by different markers. The marking scales are, however, a useful tool in many respects, not least are they a more precise description of the objectives in training writing skills.

There are a number of advantages to such analytic scales:

- they increase the validity of a test of writing as they could be regarded as representing the construct;
- they increase the validity of a test as they are a representation of course content;
- as the scales are used to assess written assignments throughout the ILSS classes, they give feedback to learners as to which areas must be improved;
- they give face validity to a test because of common assessment criteria to all learners;
- they increase the reliability of a test because of common assessment criteria for all assessors.

Of course, there are disadvantages of such an approach to assessment:

- it is time-consuming: reading a paper more than once is required considering the number of aspects/criteria;
- training is required;

- vague elements in the rating scales can be open to subjective interpretation.

In addition to these two steps – the development of marking scales and the continuous training of raters - further measures are taken to ensure the reliability of the assessment of the two writing tasks:
- student papers are not read by the class teacher but are chosen at random by the other teachers in the ILSS 2 team for marking;
- a group of second readers assesses all negative papers, but in addition they second-read positive papers as well, but do not know which of the papers they are second-marking are positive or negative;
- cases of great discrepancy between the first and second markers are read a third time and the final decision is made by the CFT co-ordinator or by all the ILSS 2 teachers at the final examiners' meeting.

The importance of continuous rater training can be demonstrated by Table 6, which shows that there are, despite all efforts to combat them, discrepancies between markers. In Table 6 the names of examiners have been replaced by letters.

Table 6: Results by markers on all three parts of the CFT – SS04 (end results compared with end results of SS03 and SS05)

Examiner	N Ss*	Read. neg.	%	Writ. 1 neg	%	Writ. 2 neg	%	Overall Neg SS04		Overall Neg SS03		Overall Neg SS05	
A	22	2	9%	4	18	5	23	5	23%	18/6	33%	23/6	28%
B	23	2	9%	5	21	1	4	2	9%	20/1	5%		
C	32	1	3%	2	6	13	41	9	28%	16/9	56%		
D	17	2	12%	3	18	4	24	3	18%	20/3	15%		
E	24	2	8%	5	21	5	21	2	8%	21/2	10%	46/10	22%
F										19/2	11%	10/0	0%
G										38/8	21%	28/8	28%
H	20	0		6	30	6	30	3	15%			22/2	9%
I	17	2	12%	4	24	6	35	5	29%			20/2	10%
J	17	1	6%	2	11	2	11	2	12%				
K												40/6	15%

* N Ss = number of students; neg = negative

Although Table 6 shows that there are stricter and more lenient markers, one should not read too much into these results without studying the composition of the student groups beforehand. However, certain conspicuous discrepancies should be and are subject of debate if the claim for standardized procedures is at all relevant.

8. General information on CFT results

CFT results are recorded as a matter of course and provide relevant data for comparison. The data referring to the CFT results can be found in the Appendices. Appendix 1, Table 1, compares CFT results to date with two pre-CFT semesters, the summer semester '02 and the winter semester '02/03. The most striking fact is the high number of drop-outs in the pre-CFT semesters and the comparatively low number of failures in these semesters. Although there are only two pre-CFT semesters with which **to** compare current data, some conclusions could be drawn: Since the introduction of the CFT, teachers use this test as the deciding factor which passes or fails a student. Whereas in previous semesters it was made clear to the student in the course of the semester or perhaps after the mid-term test that there was little chance of him/her passing the semester, the CFT is now regarded by teachers as a more objective tool of assessment which takes decision-making out of their own hands. Likewise, students might think that they will have a go at the CFT rather than leaving early if they feel they are not doing well in a course. Of course, these are assumptions, but it is worthwhile taking them into account.

Tables 2 and 3 in Appendix 2 are concerned with grading on the CFT and the final semester grades. Grading on the CFT shows a fairly similar pattern and a distinction between summer and winter semesters. The most striking features is the fact that so far there has not been a 1 ("sehr gut") grade on the CFT. There are students who receive a 1 on reading or on writing, but never on all three parts.

However, teachers have the chance of upgrading students who might have missed a 1 on the CFT, but whose course performance has been excellent. Very little use, however, is made of this option. (See Tables 3 and 5). For example, in the summer semester '05 twenty-four students received a 2 on the CFT, which could mean anything between 1.5 and 2.4, but only 2 of these were upgraded to a 1, and a further 8 of these were downgraded to a 3.

In all the semesters under consideration, the biggest changes between CFT and semester grades occur in the upgrading of 4 to 3. It is true that the CFT results are skewed towards the lower end of the grades and do not accord with the classic bell-shaped curve as can be seen on the following graphs. For that reason it is only fair that teachers have the opportunity to reward good course-work by upgrading.

9. Conclusion

With the introduction of the CFT a principled instrument for change has been put into operation with which consistency and comparability of input in the language competence classroom are furthered, and with which transparency of assessment criteria and fairness of grading are enhanced.

In this paper I have tried to shed some light on the principles behind and the development of this instrument for change. I have described how the aims and objectives of the language programme are being translated into an assessment tool, and which steps are taken to make this tool effective and efficient.

Many questions concerning the acceptability of the CFT (face validity), its overall effect on teaching (washback effect) and students' English language proficiency (validation study) must be the subject of future research, informed discussion and ultimately departmental policy. This paper has aimed to provide detailed information on the assessment procedures of interest not only to those who work in language teaching within the English department, but also to anyone interested in the transparency and comparability of the department's assessment procedures and quality management in general.

Not all statistical data available have been included in this paper as too much detail might not be of general interest. However, the most important areas of the work undertaken in connection with the design and the implementation of the CFT are recorded here. On the whole, this paper should enhance understanding of the theory and practice of standardized testing. Finally, it is hoped that colleagues who are not directly involved in ILSS teaching will take an interest in the work being done, and by adding an outsider's view, stimulate further discussion on the issues of testing.

References

Alderson, Charles J.; Clapham, Caroline; Wall, Diane. 1995: *Language Test Construction and Evaluation.* Cambridge: Cambridge University Press.

Bachman, Lyle F. 1990: *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press.

Bachman, Lyle F.; Davidson, Fred; Ryan, Katherine; Choi, Inn-Chull. 1995: *An investigation into the comparability of two tests of English as a foreign language. The Cambridge-TOEFL comparability Study.* (Studies in Language Testing 1). Cambridge: Cambridge University Press.

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

Brown, James Dean. 1988: *Understanding Research in Second Language Learning.* Cambridge: Cambridge University Press.

Clapham, Caroline. 1996: *The development of IELTS: A study of the effect of background knowledge on reading comprehension.* (Studies in Language Testing 4). Cambridge: Cambridge University Press.

Common European Framework of Reference for Languages: Learning, teaching, Davies

Davies, Alan. 2000. *Dictionary of Language Testing*. Cambridge: Cambridge University Press.

Davies, Allen. 1988: "Communicative Language Testing", in: Hughes, Arthur (ed.). *Testing English for University Study.* London: British Council, 5 – 35.

Davies, Allen; Brown, Annie; Elder Cathie; Hill, Kathryn; Lumley, Tom; McNamara, Tim. 1999: *Dictionary of Language Testing.* (Studies in Language Testing 7). Cambridge: Cambridge University Press.

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

Jordan, R.R. 1997: *English for Academic Purposes.* Cambridge: Cambridge University Press.

McNamara, Tim. 2000: *Language Testing*. Oxford: Oxford University Press.

Read, John. 2000: *Assessing Vocabulary*. Cambridge: Cambridge University Press.

Spolsky, Bernard. 1985: "What does it mean to know how to use a language?" *Language Testing* 2, 2. 180 – 191.

Spolsky, Bernard. 1995: *Measured Words*. Oxford: Oxford University Press.

Swales, John M. 1990. *Genre Analysis.* Cambridge: Cambridge University Press.

Tribble, Chris. 1996. *Writing.* Oxford: Oxford University Press.

Urquhart, Sandy; Weir, Cyril. 1998: *Reading in a Second Language: Process, Product and Practice.* London and New York: Longman.

Weigle, Sara C. 2002: *Assessing Writing.* Cambridge: Cambridge University Press.

Weir, Cyril J. 1990: *Communicative Language Testing*. Englewood Cliffs: Prentice Hall International.

Widdowson, Henry G. 1983: *Learning Purpose and Language Use*. Oxford: Oxford University Press.

Widdowson, Henry G. 2003: *Defining Issues in English Language Teaching.* Oxford: Oxford University Press.

Appendix 1: General information on CFT results

Table 1: CFT results of SS03, WS03/04, SS04, WS04/05, SS05, WS05/06 compared with the results of Sprachübung II SS002 and WS02/03

	Registration; Total	No of drop-outs	In %	No of candidates	No of negatives	In %	Drop-outs & negatives	In %
SS02	266	77	29%	189	24	12%	101	37%
WS02/03	120	21	18%	99	23	23%	44	36%
SS03	187	31	16%	156	31	20%	62	33%
WS03/04	94	5	5%	89	17	19%	22	23%
SS04	189	16	8%	173	30	17%	46	24%
WS04/05	86	7	8%	79	23	29%	30	35%
SS05	214	24	11%	190	36	19%	60	28%
WS05/06	99	10	10%	89	32	36%	42	42%

Appendix 2: Grades on the CFT

Table 2: Distribution of grades on the CFT

	SS03	156 Cand		WS03/04	89 Cand		SS04	173 Cand	
	5	31	20%	5	17	19%	5	30	17%
	4	51	32%	4	38	44%	4	53	31%
GRADE	3	52	33%	3	30	34%	3	71	41%
	2	22	15%	2	4	5%	2	19	11%
	1	0		1	0		1	0	

WS04/05	79 Cand		SS05	190		WS05/06	89 Cand	
5	23	29%	5	36	19%	5	32	36%
4	35	44%	4	71	37%	4	28	31%
3	20	25%	3	59	31%	3	27	30%
2	1	1%	2	24	13%	2	2	2%
1	0		1	0		1	0	

Table 3: Correlation between test grade and semester grade

	SS03				WS03/04				SS04			
CANDIDATES			157				86				173	
Fail	32	20%			17	20%			29	17%		
Same	77	49%	109	69%	43	50%	60	70%	94	54%	123	71%
Changes			48	30%			26	30%			50	29%
Down	10	6%			7	8%			15	9%		
Up	38	24%			19	22%			35	20%		

	WS04/05				SS05			
CANDIDATES			79				190	
Fail	23	29%			36	19%		
Same	36	46%	59	75%	102	54%	138	73%
Changes			20	25%			52	27%
Down	6	7%			14	7%		
Up	14	18%			38	20%		

Table 4: Details of downgrades

	Total	From 4 to 5	From 3 to 4	From 2 to 3	
SS03	10	2	4	4	6%
WS03/04	7	2	4	1	8%
SS04	15	3	8	4	9%
WS04/05	6	1	5	0	7%
SS05	14	1	5	8	7%

Table 5: Details of upgrades

	Total	From 5 to 4	From 4 to 3	From 3 to 2	From 2 to 1	
SS03	38	-	20	13	5	24%
WS03/04	19	-	12	6	1	22%
SS04	35	3 (pardoned)*	21	9	2	20%
WS04/05	14	2 (pardoned) *	11	3	0	18%
SS05	38	1 (pardoned)*	22	13	2	20%

* In SS04, the practice was adopted of discussing upgrades from 5-4 put forward in exceptional cases by the course teacher in the final markers' meeting where and a vote is taken on such cases.

How to contact us:

VIE*w*[z]

c/o
Institut für Anglistik & Amerikanistik der Universität Wien
Universitätscampus AAKH, Spitalgasse 2, Hof 8
A – 1090 Vienna; Austria

fax   **(intern.) 43 1 4277 9424**
eMail   **julia.isabel.huettner@univie.ac.at**
w³   http://www.univie.ac.at/Anglistik/views.htm
                      (all issues available online)