Linguistische
Arbeiten 460

Herausgegeben von Hans Altmann, Peter Blumenthal,
Hans Jürgen Heringer, Ingo Plag, Heinz Vater und Richard Wiese

The Relation of Writing to Spoken Language

Edited by
Martin Neef, Anneke Neijt and Richard Sproat

Max Niemeyer Verlag
Tübingen 2002

Die Deutsche Bibliothek – CIP-Einheitsaufnahme

The relation of writing to spoken language / éd. by Martin Neef.... – Tübingen : Niemeyer, 2002
(Linguistische Arbeiten ; 460)

ISBN 3-484-30460-X ISSN 0344-6727

Gedruckt auf alterungsbeständigem Papier.
Druck: Weihert-Druck GmbH, Darmstadt
Einband: Industriebuchbinderei Nädele, Nehren

Table of Contents

Section 4: Sharpening in German

Martin Neef, Anneke Neijt, and Richard Sproat

Introduction

This collection of papers grew out of the workshop *Writing Language*, held at the Max Planck Institute Nijmegen, the Netherlands, on August 28–30, 2000. The purpose of the workshop was to bring together researchers of diverse backgrounds who share a common goal of achieving a better understanding of the role of writing in language behavior. The international grounding of this workshop is reflected by the present volume which includes articles written by researchers working in six different countries (Belgium, Germany, Israel, the Netherlands, Switzerland, and the USA) and analyzing four different writing systems (Dutch, English, German, and Hebrew).

The papers selected for the current volume represent several lines of research into the intricate relation between writing and spoken language: Theoretical and computational linguists discuss the models that explain why orthographies are the way they are and the constraints that hold between writing and speaking a language; researchers in the area of special education deal with the question how certain aspects of orthography can be learned; and psycholinguists discuss aspects of language processing affected by variation in orthographies. Among the theoretical papers, there is one pursuing a functional perspective on language, while the others adhere to the formal paradigm, supporting either a derivational or a non-derivational theory. By offering a forum of discussion to researchers in all these fields, we hope to stimulate research that takes all aspects of the written mode into account in order to gain a better understanding of the relation between writing and the spoken language.

Several important general questions are raised by the papers to follow and we would like to review some of them briefly here.

Orthography and writing system

The terms *orthography* and *writing system* are used as near synonyms in this book. Both terms refer to the way a language is written. When a difference is intended, it will be that an orthography is the standardized set of spellings. These spellings may follow from the application of a conventional set of rules for writing a given language, or they may be singular cases that are principally independent from such a rule system. A writing system, however, includes the regularities underlying the writing behavior of competent writers which may in principle differ fundamentally from the conventional rule formulations. Most of the contributions to this volume deal with the standard spelling system, or with aspects of writing that are not explicitly standardized, in which case the more precise difference between orthography and writing system is irrelevant.

How natural is writing?

Given a modular approach to linguistic structure, one may assume that writing is a module of the grammar, with an interface level that defines the relation between the written and the spo-

ken variant of a language. Then, the relation between, e.g., orthography and phonology could be in essence comparable to the relation between phonology and syntax, and the issues discussed could be similar. Also, one may then consider writing, even though it is an artefact, to be a natural system, obeying the constraints that hold universally for the architecture of human languages.

Alternatively, writing and speaking might be considered parallel routes of processing, without a clear interface, but instead with writing being parasitic on speaking. In that case, there is no single level functioning as the interface between the oral and written modes. Given the latter point of view, one could of course also take the surface level of a language as the interface, given the assumption that language users derive all extra information to be encoded in the written mode from their knowledge of the language. And vice versa: that readers take the written form to be directly related to the language's surface structure and that extra knowledge needed to understand the written code derives from their knowledge of the language. In either case, the conclusion will be that writing is not related to the language system as if it were a natural component of the grammar.

Deep and shallow orthographies

One of the classic issues in orthographic research is the notion of *orthographic depth*. This notion is based on the ordering between the modules of the grammar, taking morphology to be 'deeper' than phonology. Within the modules, orthographic depth is based on rule ordering, classifying writing systems that encode abstract, more phonemic information as being deeper than writing systems that encode concrete, more phonetic representations. In modern models of phonology, rule ordering has been deprecated, but orthographic depth may still be a valuable notion: data do not change just because theories change, a point that is often lost in the rush to adopt new theories. The term *depth*, it seems, turns out to be a descriptive notion that is in need of a theoretical foundation and re-interpretation in actual constraint-based models of grammar.

What is the relation between orthography and the processes of writing and reading?

Theoretical linguistics aims at specifying the interrelations of the elements constituting a linguistic system, or, with regard to language users, at identifying the knowledge structures language users have to have in order to be competent. Psycholinguistics, on the other hand, deals with the question how these knowledge structures are put to use, either in production or in reception. It is an important question how these two methodological approaches are connected. Does a convincing answer in one of these fields of linguistics automatically constitute a substantive answer in the other field, or do we have to be prepared that the findings in these fields will turn out to be quite independent from each other? This problem is also relevant for research on written language. In principle, knowledge of the set of rules defining the way a language is written must be distinguished from the processes involved in applying this knowledge. Theoretical models of writing systems differ in the amount of psycholinguistic findings concerning reading and writing they are willing to incorporate. On the other hand, psycholinguistic research strongly relies on theoretical assumptions with the effect that any new trend in theoretical linguistics has strong repercussions in psycholinguistics.

Local, global, and transderivational constraints

Reflections on different kinds of constraints are suitable to further illustrate this point. Processing feasibility is one of the general constraints on language systems, and locality conditions function as baseline conditions on processing. The idea is that systems where language users need to collect information from non-local domains take too much time and effort to use, and thus cannot be psychologically real. Do such considerations apply to spelling systems as well? Recall that global constraints are constraints that refer to an earlier or later stage in the derivation. Their use led, for instance, to the introduction in syntactic theory of traces to mark the position of a moved constituent. Transderivational constraints are non-local constraints of another kind, in that they refer not only to the current derivation, but also to other, related derivations. The criticisms of non-local constraints are valid for speaking, less clearly so for writing: for example, writers may depend upon explicit instructions or conscious strategies for making the correct choice between a pair of differently spelled homophones. Writers may consciously reflect on the spelling variant needed in such cases, e.g. in choosing between the English verbs *affect* and *effect*, between *dass* and *das* in German, and between *word* and *wordt* in Dutch. One strategy for the English case, for example, is to remember that *effect* means 'to bring about'; so if one merely means 'to influence in some way', one probably wants *affect*. Such considerations seem to be non-local, since in such cases the writer seems to be invoking alternative scenarios.

So, writers consciously decide on spellings of homophones, and spelling instruction includes warnings for the writer about homophones. Furthermore, in cases of uncertainty, writers may decide to change their wording so as to avoid a potentially embarrassing mistake. However, one should not be led to the conclusion that non-local language behavior is restricted to spelling and homophony: writers also labor over matters of lexical choice (in this particular situation what is the *mot juste* ...), phrasing, morphological form (should I write *octopuses* or *octopi*), and whether, for example, a conditional is the right way to express a particular point. Even when speaking, people often think carefully about what they are saying: consider the situation where you are about to complain about something to someone and you are deciding exactly how to say it, e.g. which words and tone of voice to use, so that they won't get offended. Language use may thus include non-local processing, under special circumstances. Non-local language behavior is certainly not restricted to issues of spelling.

In speaking, however, indications of non-local behavior in language use has not led to the assumption that non-local constraints are available for the lexicon, morphology, syntax, and semantics. Rather, one of the basic assumptions has been that language systems are constrained by severe locality conditions, excluding global and transderivational constraints from the description of language. Similar considerations may be taken as point of departure for spelling research, but the facts that writing language is a more conscious process and learning to write requires explicit instructions may give us some indication that writing systems are essentially different from natural language, and that these may exhibit non-locality, such as global or transderivational constraints.

The dependency hypothesis versus the autonomy hypothesis

Theories of orthography usually follow a conception that seeks to derive written forms from spoken forms. This is most obvious for the relation between sounds and letters. The sounds

represented in phonological structures are taken as the primary elements on which the respective letters are dependent. Under this view, the English word *beat* has the initial letter <b> because the underlying spoken form [bi:t] begins with the sound [b]. However convincing this approach is on first sight, there are several aspects of written forms that cannot be explained straightforwardly in this way. For example, there are letters that have no basis in the pronunciation, as in the case of the mute <h> in German, and there are sounds that have no reflection in the spelling, as in the case of short vowels in unvocalized Semitic writing systems. The use of graphotactic constraints in Dutch further illustrates autonomy: there is no phonological difference between <a> and <aa> in *manen* 'moons' and *maan* 'moon', and the difference can be described with rules that refer to the string of letters only. It will be an issue of future research to decide on the balance between the dependency hypothesis that highlights aspects of spellings that have a clear base in the spoken forms or the autonomy hypothesis which focuses on those elements of spellings that seem to have a status independent from the spoken forms. When both kinds of rules are needed for a proper understanding of writing systems, three sets of information about a given writing system are implied: Well-formedness constraints on the output (the strings of letters, the use of spaces, punctuation, and perhaps layout); rules governing the relation between the spoken language and its written output; and rules governing the opposite relation, between writing and spoken language.

Readability versus writability

This theoretical dichotomy can be subsumed under a more general view on orthography: What is the balance between reading and writing in orthography? At the design phase of a writing system, the need to express what can be spoken is present, but in the case of spelling reforms, the needs of the readers may become more important. This may explain why spaces in between words were invented relatively late. Perhaps also the general tendency of spelling systems to develop from more phonologically based to more morphologically based can be explained this way. Approaches to explain orthographies predominantly stem from the perspective of the writer. This is understandable given that learning to write is much more difficult than learning to read. Hence, the didactics of orthography are the didactics of writing. Efforts to reform a specific orthography also predominantly stem from the perspective of the writer. This may be because lecturers in teaching methods have been given the main responsibility of spelling reforms. But it may lead in a wrong direction, if it turns out that the main function of a writing system is not to make writing as easy as possible but to make reading as effective as possible. Thus, the question of readability may be one of the central aspects of future research on writing systems.

The contributions in this volume

The contributions to this volume all deal with one or more of the tenuous questions posed above. The first section is devoted to the discussion of a theoretical conception introduced by Sproat (2000), the *Consistency Hypothesis*. Embedded in a derivational conception of grammar, this approach makes the substantive claim that for each language there is one fixed point in the grammatical derivation where the derivation of the writing system of that language branches off. Sproat terms this point in the derivation the *Orthographically Relevant Level*. As a consequence, the effects of some linguistic rules should be consistently mirrored in the

respective written forms while the effects of other rules should consistently not be visible in the written forms. In her paper *The Interfaces of Writing and Grammar*, Anneke Neijt challenges the Consistency Hypothesis. On the basis of Dutch, she claims that only the first step in the translation of sounds into letters can be restricted to one consistent level. For the graphotactic rules defining the well-formedness of strings of letters and for other aspects of written forms, more than this single level is necessary in defining the relation between speaking and writing. Among these other aspects is punctuation that is in need of global information from morphology, syntax, and semantics. Furthermore, morphological information has to be invoked. Classes of morphemes may form exceptions to an otherwise consistent spelling system, with depth in terms of their derivations having no bearing on the issue.

In *The Consistency of the Orthographically Relevant Level in Dutch*, Richard Sproat carefully examines the data presented by Neijt and concludes, that given certain assumptions about rule formulations, a specific phonological level (a level somewhere in between phonemes and phones) can be taken as the input of the writing system for a language such as Dutch. Sproat supplies an explicit analysis of a fragment of Dutch phonology that gives a clear localization of the Orthographically Relevant Level in Dutch, dealing with questions like stress, final devoicing, and different rules for native vs. non-native morphemes. In his conclusion, Sproat reflects on the naturalness of the Consistency Hypothesis.

The second section presents cross-linguistic studies. Susanne Borgwaldt and Annette de Groot base their paper *Beyond the Rime: Measuring the Consistency of Monosyllabic and Polysyllabic Words* on a close inspection of the writing systems of Dutch, English, and German. Their focus is the notion of consistency in a psycholinguistic tradition. Usually, research on phonological consistency focuses on monosyllabic words, which are split up into onset and rime. Subsequently, the mappings between written and spoken rimes are compared. Words sharing the same written rime are then considered feedforward consistent if the corresponding spoken rimes are pronounced in the same way. Words sharing the same spoken rime are called feedback consistent if their rimes are written in the same way. Borgwaldt and de Groot offer a method for determining the degree of bidirectional consistency that is applicable for monosyllabic and polysyllabic data alike. It is shown that by taking not only the consistency mappings between rimes into account but also those between other (overlapping) subsyllabic units, the accuracy of the description of consistency increases considerably.

In *Teachers' Perception of Spelling Patterns and Children's Spelling Errors: A Cross-Linguistic Perspective*, Dorit Ravid and Steven Gillis illustrate that the complexity of orthographies as different as Hebrew and Dutch must be valued from different perspectives. They examine the teachers' perception of morphologically-mediated spelling patterns, compared with children's actual spelling performance on items spelled according to these same patterns. The study focuses on teachers' explicit knowledge of the role of morphological and morphophonological cues in spelling homophonous graphemes in Hebrew and Dutch, with alternative spellings for the same sound. In general, Ravid and Gillis find that teachers' metalinguistic knowledge of spelling patterns is a mirror image of children's performance. The authors explain their findings in terms of consciousness: explicit metalinguistic formulation of spelling patterns operates differently than natural information processing in language use.

Section 3 deals with elements of writing systems different from mere letters. One kind of such elements are diacritics that modify the content of letters. A specific type of diacritics is

the topic of *Effects of Diaeresis on Visual Word Recognition in Dutch* by Vincent van Heuven. Usually, efforts of the writer make the reading process easier. For instance: when writers bother to signal nouns with capital letters (as in German), the reading process will be facilitated. This is not what has been found by Van Heuven for the use of diaereses in Dutch. Such diaereses signal orthographic syllable boundaries and sometimes prevent homography. In a lexical decision task manipulating words with and without diaereses and with a transposed diaeresis, however, reaction times were not faster, nor was the accuracy of lexical decision different. This shows that not all information on the pronunciation of words need to be encoded in the written form. A certain amount of abstractness will lead to a system which is more efficient for the writer and nonetheless equally efficient for the reader.

Punctuation marks have a less clear foundation in spoken language than letters. The theoretical debate revolves around the question to what amount their distribution can nevertheless be explained with recourse to the phonological structure of linguistic units. Jochen Geilfuß-Wolfgang in his article *Optimal Hyphenation* intends to find supporting evidence for an autonomous approach to orthography. According to his analysis, hyphenation is sensitive to orthographic syllables, a notion that is related to, but not identical with, the phonological syllable. Employing the constraint-based Optimality Theory, Geilfuß-Wolfgang formulates some constraints specific to the orthographic component of grammar. If these constraints are adequately ranked, the account enables the computation of the hyphenation data in German. Geilfuß-Wolfgang concludes that orthographic syllables and phonological syllables have many of the same properties and are governed by many of the same structural constraints.

Other punctuation marks like the comma or the full stop cannot be explained in relation to word phonology. As Ursula Bredel in *The Dash in German* shows, analyses that assume a dependence of written forms on spoken forms argue whether intonation or syntax primarily guide the distribution of punctuation marks. Since intonation itself is grounded in syntax, however, these approaches can both be regarded as syntactical in nature. Bredel herself opts for a different approach that focuses on the characteristics of written language and can, thus, be subsumed under the autonomy paradigm. In general, she takes punctuation marks as means for the steering of language processing. Based on a historical reconstruction of the functions the dash has had in German orthography, Bredel is able to reduce the diverse manners of use of the dash in contemporary German to one main function, namely to prepare the reader for a shift of focus.

Section 4, the final section of the book gives different perspectives of one particular noteworthy phenomenon of the writing system of German, namely *sharpening*, which is also known as *consonant doubling*. The core of this subject can be illustrated by the word *Neffe* 'nephew': Its pronunciation [nɛfə] contains only one fricative, but the corresponding letter <f> appears twice in the written form. Many different approaches have been supported to explain this complex. Christina Noack compares three different rule systems to sharpening, spanning a period of more than two hundred years, in her paper *Regularities in German Orthography: A Computer-Based Comparison of Different Approaches to Sharpening*. These rule systems differ in their central focus, which is either the segment or the syllable or the morpheme. Noack's main concern is to present a computational tool to evaluate the consistency of alternative linguistic analyses on the basis of large corpora. This computer program, called ORTHO 3.0, enables her to give explicit lists of exceptions that each of the rule sys-

tems generates. With respect to the number of exceptions, the segment-based approach shows the worst results, while the other two approaches do not differ significantly.

The approaches compared by Noack all share the dependency perspective in that they derive the written forms from the spoken forms. Martin Neef in *The Reader's View: Sharpening in German* offers an alternative analysis within the autonomy paradigm. He follows the idea that the function of orthography is to give the reader instructions on how to read an unknown text. The basic tenet of this approach is therefore the *Readability Principle*, which demands that spellings should guarantee an unambiguous access to spoken forms. On this background, Neef re-examines the sharpening-data. His analysis reveals slightly different exceptions from the derivational theories discussed in Noack's article. Furthermore, Neef uncovers areas of the vocabulary that show orthographic underspecification, and he defends a different position on the question in how far sharpening is stress-based.

The final paper of the volume, *How Syllable Structure affects Spelling: A Case Study in Swiss German Syllabification* by Thomas Lindauer, takes up the theme of dialectal variation. It is well-known that language communities are non-homogeneous. This holds especially for the German speaking societies in Germany, Switzerland, and Austria. Nevertheless, one writing system is agreed upon for these non-homogeneous groups of speakers. The question arises, then, how this writing system can be taught most effectively, given that explicit learning rules rely on phonological awareness of the learners. Most difficult are, of course, those rules that refer to phonological information not available for a group of language users (because a certain distinction is lacking in this variant of German). Lindauer proposes to present explicitly different spelling rules for the different communities, nevertheless leading to the same spelling output. He illustrates his assumptions with the example of sharpening and the related phonological phenomenon of ambisyllabicity. Since the phonological structures related with ambisyllabicity are different in Standard German and in Swiss German, the rules teaching sharpening should be different for these language communities.

Most of the papers presented here grew out of oral presentations at the workshop *Writing Language*. Other talks of the same workshop will be published separately in an issue of the *Journal of Written Language and Literacy*, edited by Rob Schreuder and Ludo Verhoeven. The workshop was organized by Harald Baayen, Martin Neef, Anneke Neijt, Rob Schreuder and Ludo Verhoeven, and sponsored by the Nederlandse Organisatie voor Wetenschappelijk onderzoek (NWO) and the Center for Language Studies (CLS). We greatly appreciate this support that helped in producing this book. Finally, we would like to thank Richard Wiese, the editor of *Linguistische Arbeiten*, for many helpful comments, and Moritz Neugebauer and Jessica Schwamb (University of Cologne, German Department) for their help during the final stages of the preparation of this book.

Section 1: Consistency

Anneke Neijt

The Interfaces of Writing and Grammar

1. Introduction

Spoken language, sign language, and written language – three modes of expression, but one underlying system? The answer will be negative for sign languages. Studies reveal that sign languages need not be derived from spoken languages and that if they happen to be derived from spoken languages, they tend to develop characteristics not present in their spoken origins (Wilbur 1987, Boyes Braem 1995). For younger generations, sign language can be acquired in a way that is familiar to how spoken languages are learned. Hence, there is evidence that sign language forms a system on a par with spoken language and is not dependent on it.

This is not the case for the written mode. Writing seems to be secondary to oral language, being derived from it, and fundamentally different from sign language. Each new generation learns the written variety at school after most of the spoken language has been acquired. Whereas for children the acquisition of a spoken or sign language is an unconscious process, acquisition of writing requires explicit learning strategies, of which teachers and pupils are well aware. Writing should be considered another code for the language acquired, which is why spelling is called secondary. The existence of spelling pronunciations, however, shows that this secondary mode of expression influences speaking, the primary mode (Van Haeringen 1962, Wells 1982: 106-9, Carney 1994, Maas 2000: 33). Other evidence for this influence on the primary mode comes from psycholinguistic experiments (cf., for instance, Seidenberg & Tanenhaus 1979, Schreuder et al. 1998) and from language change (Jespersen 1909). In this paper, the question how both modes of expression are related is investigated from a theoretical point of view.

The close relationship between a spoken language and its written variant has led to the hypothesis that the major part of the system is shared by both modes of expression. For instance, the semantic component provides the interpretation of scope-bearing elements, whether written or spoken; the syntactic component provides word order for both. Morphology creates words and inflection for both, and even some part of phonology is common, e.g. phonological segments correlate closely with letters. Some writing systems are called 'deeper' and others more 'shallow', reflecting the derivational level relevant for writing. Systems based on morphosyntactic structure are called deeper than systems based on phonological or phonetic representations (Haas 1976, Sampson 1985, Sgall 1987, Asher & Simpson 1994, Daniels & Bright 1996, Meisenburg 1996). The claim is that the written mode of expression follows a route different from the oral mode only in the final stage of processing. In reading, it is only the first stage of processing that follows a different route, according to this hypothesis. Speaking and writing thus share a large number of derivational stages, as do hearing and reading. Schematically:

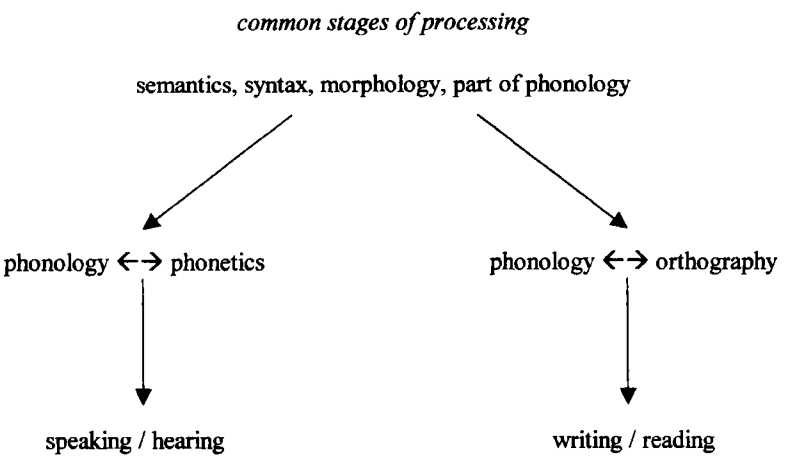

Figure 1: *General model of the relation between spoken and written language*

This view on how spoken and written language relate to each other has been worked out for Dutch by Nunn (1998). Dutch orthography is known to be based on a deep phonological stage of processing, cf. Van Heuven (1978) and Booij (1987). Nunn (1998) adds to this the conclusion that the derivation from phonology to orthography consists of two steps. After the first step of phoneme-to-grapheme conversion for morphemes, a second step takes care of grapheme co-occurrence restrictions by way of graphotactic rules, i.e. grapheme-to-grapheme conversion rules. Nunn calls such rules 'autonomous spelling rules', claiming that the rules refer to orthographic information only, although some of the phonological characteristics (the distinction between consonants and vowels, for instance) are carried over to the orthographical representation.

Of course, in defending the claim of a derivation in two steps, Nunn emphasizes the differences between the two steps, i.e. the difference between phonologically and orthographically based rules. It is from this perspective that Nunn tries to find evidence for the orthographic nature of autonomous spelling rules and to restrict the amount of phonemic information necessary for the second step in the derivation from phonology to writing. From this perspective, it is not surprising that Nunn's analysis of Dutch has been used in Sproat (2000: 16) to illustrate the Consistency Hypothesis.

(1) *Consistency*
The Orthographically Relevant Level for a given writing system (as used for a particular language) represents a consistent level of linguistic representation.

This hypothesis, a direct reflection and strict interpretation of the model sketched in figure 1, states that there is one consistent Orthographically Relevant Level for a given writing system, not more than one, cf. figure 2. Notice that 'Consistent' here must not be understood as 'without exceptions'. Where alphabetic writing systems concern the spelling of finite sets of elements, the opportunity is present to store exceptional orthographic forms in memory. It seems that exceptions occur in many alphabetic writing systems.

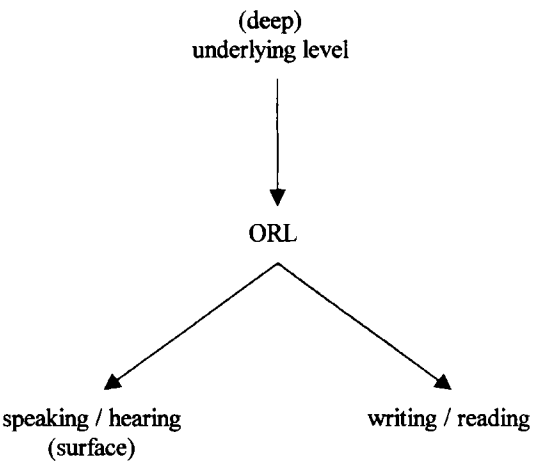

Figure 2: *The claims of the Consistency Hypothesis: one consistent level by the oral and written modes*

In this paper, evidence will be presented to show that the processes of speaking and writing share more information than can be provided by a single derivational level. The claim made in this paper is that the phoneme-to-grapheme conversion rules are based on information from different levels, as are the grapheme-to-grapheme conversion rules. Of course, the distinction between the two sets of rules will be valid even when more than just one linguistic level provides input to the orthographic representation. Therefore, the two-step analysis of Nunn can be maintained, though defined in a less rigorous fashion. The Consistency Hypothesis, however, cannot be maintained as a universal principle.

The layout of this paper is as follows. First, the arguments by Nunn (1998) in favor of a two-step derivation of orthography will be reviewed. Then, in section 3, the Orthographically Relevant Level according to Nunn will be discussed. It will be shown that the hypothesis that there is only one such level can be maintained only at the cost of storage. Sections 4 and 5 show that a native Orthographically Relevant Level must be distinguished from a non-native Orthographically Relevant Level and that punctuation is based on other levels than the phonemic representation of morphemes. Section 6 presents the linguistic information necessary for the autonomous spelling rules. Section 7 finally summarizes the evidence gathered in the preceding sections about the linguistic levels needed for writing and presents the overall conclusion. Information from different levels of language processing is collected in writing. In the presentation that follows, most arguments are based on writing and virtually no arguments are presented about reading.

2. An outline of Dutch orthography

Detailed information on the orthography of Dutch can be found in Nunn (1998). She distinguishes several orthographic components that are relevant for Dutch. Conversion of native morphemes needs to be distinguished from conversion of non-native morphemes, and a set of autonomous rules forms part of the orthographic derivation. Figure 3 is Nunn's analysis in a nutshell. Observe that she assumes one level with information on the underlying, phonemic, representations of the segments of morphemes at which all information of the spoken mode is translated into information on the written mode. Nunn's proposal for Dutch therefore confirms Sproat's Consistency Hypothesis. According to Nunn, there is one Orthographically Relevant Level, the level of morphemes in their phonemic form:

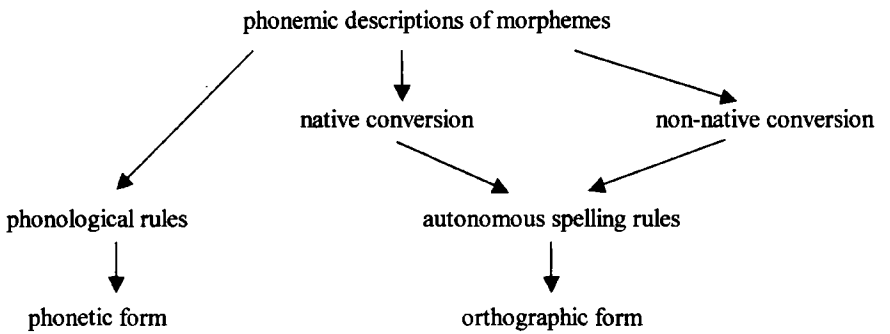

Figure 3: *Nunn's model of the relation between phonetic and orthographic form*

The remainder of this section will present explanatory notes on this model.

Dutch has a so-called deep orthography. Underlying rather than superficial sound segments are spelled; i.e., morphemes tend to receive a uniform spelling, irrespective of the application of certain phonological rules that generate sets of allomorphs. Frequently used examples to illustrate this are *hond* and *heb*, with final obstruents spelled in accordance with their underlying forms /hɔnd/ and /hɛb/ instead of their phonetic forms [hɔnt] and [hɛp]. These underlying forms are detectable for the writer on the basis of plural inflection: [hɔndə] and [hɛbə] with voiced obstruents. Other examples are *zuinigheid*, *aanmelden*, *hoofddoek* 'carefulness, to announce, head-shawl', for which a more superficial spelling would be **zuinigeit*, **aamelde*, **hoofdoek*, derived by h-deletion, final devoicing, nasal assimilation, final n-deletion, and degemination.

Furthermore, Dutch is a language with two sets of words: native ones, such as *kunstzinnigheid*, and non-native ones, such as *artisticiteit*, both meaning 'artisticity'. The difference has its origins in the earlier stages at which Dutch imported words from Latin or French, but new borrowings follow this distinction as well. Non-native words can be distinguished from native ones on the basis of systematic differences in present-day phonology and morphology (Van Heuven et al. 1994, Nunn 1998: 155 ff.). One of the most important characteristics is the

number of full vowels present in morphemes: when more than one full vowel is present, the morpheme will be non-native. Exceptions to this rule are only a handful of frozen compounds such as *aardbei* 'strawberry' which behave as native words, notwithstanding the presence of more than one full vowel.

The distinction between native and non-native morphemes takes the native morphemes as point of departure, such that all morphemes not in accordance with the constraints that hold for native morphemes are non-native. Therefore, the fact that only one full vowel is present in a morpheme is a necessary but not a sufficient criterion for this morpheme being a native morpheme. Further constraints are the combination of consonant clusters (for instance, only a limited set of clusters occurs in native morphemes, not the clusters /sk/, /sf/, and /tm/, which predicts that *skelet* 'skeleton', *sfeer* 'sphere', and *ritme* 'rhythm' are non-native words, even though only one full vowel occurs) and constraints on morphology (for instance: plural -s is restricted to native words ending in /a, o, u/ and native words ending in a syllable with schwa; hence, the plural forms *trams* and *e-mails* indicate that these words are non-native). On the basis of such criteria, the etymological distinctions are recoverable from the synchronic spoken mode even for language users without any knowledge of foreign languages.

The orthography reflects the difference between native and non-native words, since partly different sets of phoneme-to-grapheme conversion rules are used (indicated in figure 3 by the two routes for native and non-native morphemes) with, for instance, the graphemes *c*, *q*, *th*, *y*, and *x* for non-native words only, cf.:

(2)	*non-native*	*native*	*sounds*
	camera 'camera'	kamer 'room'	/k/
	quasi 'quasi'	kwaad 'angry'	/k/
	ether 'ether'	eter 'eater'	/t/
	hypothese 'hypothesis'	hier 'here'	/i/
	examen 'test'	heks 'witch'	/ks/

Literacy therefore leads to awareness of the distinction between native and non-native morphemes.

The general model in figure 1 of how speaking and writing can be related is not only complicated by the difference between the spelling of native and non-native words, but also by the existence of autonomous spelling rules. One of the reasons to incorporate such rules in the model of Dutch orthography is the presence of allography in examples such as:

(3)	*stem*	*derived form*		*spelling*	
	bak	– bak+er	→	bakker	'baker'
	judo	– hij judo+t	→	hij judoot	(third person ending of the verbal stem *to judo*)
	laan	– laan+en	→	lanen	'lanes'
	vers	– iets vers+s	→	iets vers	'something fresh'

No phonological alternation is involved here. In order to account for such forms of allography, Nunn (1998: 183 ff.) proposes a set of autonomous graphotactic rules, i.e. rules that operate on grapheme sequences, such as the following ones for gemination and degemination. The formulation of Nunn's rules has been simplified for expository reasons. C abbreviates for consonant letters, V for vowel letters, and dots indicate syllable boundaries. The distinction between short and long vowels is not one of phonetic duration, but rather expresses the fact

that short vowels may combine with a coda that consists of more consonants than the coda following long vowels.

(4) a. *Orthographic gemination*
C → CC after a short vowel at the end of the syllable
V → VV for long vowels when a C follows within the syllable
b. *Orthographic degemination*
VV → V when syllable final
CC → C when syllable final

The derivation of the words presented in (4) runs as follows (backslashes indicating the underlying orthographic forms):

(5) a. *Conversion of morphemes*

/bɑk/	→	\bak\	/ər/	→	\er\
/jydo/	→	\judo\	/t/	→	\t\
/lan/	→	\laan\	/ən/	→	\en\
/vɛrs/	→	\vers\	/s/	→	\s\

b. *Concatenation of morphemes and syllabification*
\ba.ker\
\ju.dot\
\laa.nen\
\verss\
c. *Application of orthographic (de)gemination rules, cf. (4)*
<bakker>
<judoot>
<lanen>
<vers>

As a result of these orthographic rules, vowel letters for short vowels are always followed by a consonant within the syllable, whereas syllable-final vowel letters represent long vowels. It is because of this pattern that short and long vowels in the literature on Dutch orthography are called 'covered vowels' and 'free/uncovered vowels' (Dutch *gedekte* and *ongedekte/vrije vocalen*). Covered vowels are always followed by a consonant letter within the syllable, whereas uncovered vowels may occur at the end of syllables:

(6)

covered/short vowels		*covered by C-gemination*		*uncovered/long vowels*	
[kɑntə]	kan.ten 'sides'	[kɑ.nə]	kan.nen 'cans'	[manə]	ma.nen 'moons'
[kɛldər]	kel.der 'cellar'	[bɛ.lə]	bel.len 'bells'	[benə]	be.nen 'legs'
[plɔftə]	plof.te 'plumped'	[plɔ.fə]	plof.fen 'to plump'	[pokər]	po.ker 'poker'

This generalization holds in orthography but is present in phonology as well: intervocalic consonants after short vowels are ambisyllabic, as demonstrated in experiments in which speakers of Dutch are forced to explicitly syllabify such examples (cf. Rietveld 1983 and Sandra et al. 1996). The experiments show that speakers' judgments are influenced by orthography. However, interestingly, illiterate speakers of Dutch and pre-school children also present analyses with ambisyllabic consonants, though significantly less than the literate participants for whom the spelling rules seem to enhance ambisyllabic responses.