Hans-Christoph Nürk

On the Importance of Visual Determinants in Visual Word Recognition

DISSERTATION

Philipps-Universität Marburg

On the Importance of Visual Determinants in Visual Word Recognition

Dissertation

zur

Erlangung eines Doktorgrades

der Naturwissenschaften

(Dr. rer. nat.)

dem

Fachbereich Psychologie

der Philipps-Universität Marburg

vorgelegt von

Hans-Christoph Nürk

aus Göppingen

Marburg / Lahn, 2001

Vom Fachbereich Psychologie

der Philipps-Universität Marburg als Dissertation

am

angenommen.

Erstgutachter: Prof. A. M. Jacobs

Zweitgutachter: Prof. H.-H. Schulze

Tag der mündlichen Prüfung am

Wir leben von dem, was uns Menschen in bedeutungsvollen

Stunden unseres Leben gegeben haben.

(Novalis; aus der Anzeige zum Tode meines Vaters,

12.7.1935 – 7.2.2001)

We live from what people have been giving us

in meaningful hours of our life.

(Novalis; from the obituary notice of my father's death,

7/12/1935 – 2/7/2001)

Abstract

This thesis is concerned with the importance of visual determinants in visual word recognition and its possible underestimation due to inappropriate models and designs. The issue is examined in four respects. First, the theoretical foundations of a possible underestimation are introduced. It is suggested that the application of theoretical and statistical models and principles may not always have been appropriate. Second, empirically, it is shown that main and interaction null effects of visual determinants are not reliable because such effects can be detected with appropriate manipulations, presentation conditions, and analyses, but may be missed otherwise. Third, fits using the model of Nazir, Jacobs, and O'Regan (1998) suggest throughout this thesis that a linear, non-interacting account of visual and lexical factors (e.g., in the general linear model) is not appropriate. Finally, a nonlinear arithmetical model of word recognition (NAMWR) is introduced in which the interaction between visual and other variables is incorporated in a non-linear way in the specific domain of (visual) word recognition. This model provides successful fits to the data. Thus, it presents an alternative theoretical account and strengthens the notion of inappropriate application of other theoretical models.

The role of visual determinants was investigated in the above four respects in three different domains. First, it was shown that perceptual learning plays an important role in the reading process. The item-specific perceptual frequency of the visual word form determined performance in Study 1. Words tended to be best recognized in that particular form, in which they are most frequently encountered. In addition, it was shown that length-specifically words tended to be perceived best on that viewing position on which they are most frequently encountered. Second, Study 2 investigated the frequency effect and its interactions with other variables under different visual presentation conditions. The (near-optimal) visual presentation conditions used in most laboratory experiments most likely produce null effects. Third, Study 3 showed that the orthographic-lexical relations between words are not only determined by the orthographic structure of the lexicon. In two tasks, the neighborhood size effect varied considerably with the visual confusability of the target words with their neighbors. Therefore, neighborhood effects may be better understood as similarity indices in visual-orthographic-lexical space: Their investigation under one certain visual condition is thus not sufficient to form a general conclusion that is independent of visual determinants.

It is concluded that the importance of visual determinants should not be underestimated because even hallmark effects of visual word recognition research (frequency, neighborhood size, and word length effects) can change as a function of visual determinants. The NAMWR represents a new arithmetic model that numerically specifies how the non-linear alteration of these hallmark effects can be successfully fitted and possibly better understood than with the general linear model using additive factor logic.

Table of Contents

I. General Introduction: On the Importance of Visual Determinants in Word Recognition

1. Introduction

The limitations of given models provide the impetus for the next generation of research.

(Seidenberg & Plaut, 1998, p. 236).

The purpose of this thesis is to investigate the importance of visual determinants of visual word recognition. This importance has sometimes been rejected based on statistical null effects in the general linear model. However, I will argue that these null effects do not necessarily imply that visual aspects do not have much effect on visual word recognition performance. Instead, such null effects may also be due to the limitations or inadequacy of statistical models or experimental principles. This thesis will try to provide evidence that with appropriate stimulus and presentation manipulations one can discover previously neglected visual effects and visual modulations of standard effects. Based on these findings, it will be discussed which implicit theoretical assumptions in the experimental method or the underlying models used in other studies led to such rejections of visual effects. Finally, a mathematical model will be proposed that accounts for most effects found here. The model will numerically specify assumptions that differ from the theoretical assumptions found to be inappropriate in some of my previous experiments.

At the beginning of this introduction, I will give a short overview over perceptual frequency effects in other visual domains than visual word recognition. As an example for a perceptual frequency effect in visual word recognition, I will introduce the viewing position effect and the mathematical model of Nazir et al. (1998) that can account for this effect. This model and in particular its current limitations (see quote above) will help to specify hypotheses in the first experiments of each of the three studies in this thesis. These experiments are guided by the hypothesis that perceptual frequency does not only extend to a perceptual learning hypothesis of location (Nazir et al., 1998) but also to a perceptual learning hypothesis of item-specific visual word form. However, this hypothesis is in contradiction with some part of the current literature. Therefore, I need to review this literature. I will try to point out that the theoretical assumptions underlying the notion that item-specific visual word form does not play any role in visual word recognition (e.g., Besner, 1989) may not be appropriate. These inappropriate theoretical assumptions refer mainly to the ideas that i) the general linear model used for ANOVA design is generally assumed to be true, and that ii) based on the general linear model null interactions can be interpreted

according to additive factor logic (AFL) as evidence for independent additive stages and against interactive models in particular (Sternberg, 1969, for AFL and an early warning against such an interpretation). This criticism lays the ground for Study 1 in which a previously neglected visual process, perceptual learning of item-specific word form, is investigated. However, the possible inappropriateness of the underlying theoretical assumptions is also important for the other two studies which examine interactions of visual determinants with hallmark effects of visual word recognition.

In Study 2 I am not concerned with a rediscovery of previously neglected visual effects in visual word recognition but rather with the alteration of the frequency effect by visual factors: If a significant interaction of frequency and visual manipulations has been found, then the frequency effect has quite consistently been shown to increase under non-optimal or poor visual conditions. This is not only true for the frequency effect itself but also for the interaction of frequency with other variables, such as word length in Study 2. In a similar way as for the perceptual frequency hypothesis of visual word form, a rejection of an alteration of the frequency effect as a function of other visual variables as well as a neglect of an interaction of frequency and word length may due to an inappropriate application of AFL. Moreover, the principle of isolated variation which assumes that the effects of one manipulation can be studied when all other factors are held constant on a certain level may not be applicable here because not only the frequency effect itself varies with visual conditions but also the interactions of frequency with other variables do. These interactions seem particularly to emerge under poor visual conditions. Thus, null interactions of frequency with other variables observed under constant near-optimal visual conditions in laboratory experiments can be misleading. The interaction I will investigate is the interaction between frequency and word length. Like the frequency effect, the word length effect has also been shown to increase under poorer visual conditions. As I am interested in an interaction of frequency and word length, I will also give a short review of how word length interacts with other variables. However, the main focus will be on the interactions with word frequency and with variations of visual presentation conditions.

I will test the hypothesis that inferior visual conditions enhance the frequency effect, the word length effect, and their interaction by manipulating viewing positions. However, I do not only want to analyze the empirical results in an ANOVA but also want to interpret them within the framework of the model of Nazir et al. (1998). Because the model of Nazir et al. (1998) can currently only fit viewing position data for a fixed number of letters and fixation positions, I will - in a first necessary refinement - generalize the model to different word lengths and any number of viewing positions. This refinement will allow us to gain more insight into the interaction of word length with frequency on different viewing positions because the

specific nature of the interaction can be better interpreted than by simply analyzing interaction patterns in ANOVAs. The fitting results will then be used as a constraint for a nonlinear arithmetical model of word recognition (NAMWR) at the end of this thesis.

In the final Study 3, I will take a step beyond individual properties of a word and study the nature of the alteration of its relations to other words in orthographic-lexical similarity space. Therefore, I will take a closer look on certain orthographic-lexical effects in visual word recognition, the neighborhood effects (Grainger & Jacobs, 1996). Neighborhood effects are mostly regarded as pure orthographic-(phonological)-lexical effects that can be studied according to the principle of isolated variation on central fixation with unlimited presentation. Again, this design and theoretical idea may possibly not be applicable in some cases because neighborhood may interact in a characteristic way with visual variables, namely the legibility and confusability of the critical letters for neighbor creation. I shortly review neighborhood effects and what is known about their alteration by visual presentation conditions. On this basis, I will try to raise doubts whether neighborhood effects can be studied independently from visual variables. Instead, it will be hypothesized that the investigation of neighborhood under near-optimal conditions, i.e. unlimited central fixation, may produce results that are specific to those presentation conditions and cannot be generalized. Again, this hypothesis can be further specified within the model of Nazir et al. (1998). I will examine which parameters of the model should be affected or not affected when the distribution of the letter positions of neighbors (neighborhood distribution) and viewing position are manipulated.

On the basis of the results in the three studies and the constraints imposed by the model fits of Nazir et al. (1998), a new mathematical model of the viewing position effect, the frequency effect, neighborhood effects, the word length effect, and their interactions with one another will be proposed (the NAMWR). The specific incorporations of this model represent an alternative account of the nature of these effects and their interactions that contrasts with AFL and general linear model assumptions criticized in the introduction.

2. On the Importance of Visual Determinants in Visual Word Recognition

2.1. Introduction

Reading uses visually based letter clusters of the same size and case

(Mayall, Humphreys, & Olsen, 1997, p. 1285).

In this thesis I will try to show that this assumption which is based on thorough investigations of case-mixing effects in different reading conditions, can be subsumed under a more general hypothesis: The perceptual frequency hypothesis of item-specific visual word form. It simply states that:

The more often a word is recognized in a particular item-specific visual form, the better it will be recognized when presented again in that particular visual form.

This perceptual frequency hypothesis implies two controversial assumptions:

1. Perceptual (and not only lexical) learning takes place every time a word is encountered.
2. Perceptual learning is not restricted to general perceptual units (like single letters encoded in abstract form, Evett & Humphreys, 1981), but is also item-specific.

Clearly, for common English words this perceptual frequency hypothesis leads to the same predictions as other accounts of visual word form effects: It implies that English words should best be recognized in same case and same size (and not in mixed case). Indeed, this effect has frequently been found and replicated (Smith, 1969; Smith, Lott, & Cronell, 1969; for reviews see Mayall et al., 1997; Paap, Newsome, & Noel, 1984). The perceptual frequency hypothesis also predicts that under normal perceptual circumstances English words should best be recognized in lower-case and in a typical print type. This also seems to be the case (for reviews see, for example, Brown & Carr, 1993; Paap, et al., 1984). However, these effects may be obtained because item-specific perceptual learning occurs and not only because of general attributes of mixed case or lower-case stimuli.

In this introduction, I will first summarize evidence for perceptual learning effects in vision and visual word recognition and briefly discuss a mathematical model that can account for one such effect, the viewing position effect (VPE). Then an overview of studies concerned with item-specific visual word form processing will be given and *'The perceptual frequency hypothesis of item-specific visual word form'* will be discussed in greater detail. Finally, methodological and theoretical objections will be raised against

interpreting null effects in the general linear model as conclusive evidence against perceptual learning effects.

2.2. Perceptual Learning in Other Visual Domains

In recent years, visual perceptual learning has been demonstrated in many psychophysical tasks for different stimulus attributes such as motion direction (e.g., De Luca & Fahle, 1999; Fahle & De Luca, 1994; Vaina, Sundareswaran, & Harris, 1995), stimulus orientation (e.g., Ahissar & Hochstein, 1993, 1996, 1997; Fahle & Edelmann, 1993), or retinal position (e.g., Ahissar & Hochstein, 1996; Karni & Sagi, 1991; for recent summaries see Ahissar, Laiwand, Kozminsky, & Hochstein, 1998; De Luca & Fahle, 1999). In all these studies (extensive) practice of simple visual tasks lead to significant performance improvements in human adults. However, mostly, this improvement was restricted to the specific properties of the stimulus and did not or rarely transfer or generalize to, for example, different orientations or motion directions (Ahissar & Hochstein, 1997; Ahissar et al., 1998; De Luca & Fahle, 1999). For the purpose of this thesis the important points are that a) perceptual learning of specific configurations and items is possible, and b) perceptual learning can occur at various levels of the visual system (Ahissar & Hochstein, 1997) and therefore possibly for different visual unit sizes. The latter point converges nicely with the conclusion of Mayall et al. (1997) that their *"results favor a model of visual word processing in which lexical access is based on units coded at multiple levels: Single letters, transletter features, and familiar letter groups"(p. 1285).* However, Mayall et al. (1997) deny the possibility that more complex visual patterns than familiar letter groups such as item-specific word forms can also be perceptually learned (see also Besner & Johnston, 1989). This thesis will investigate whether perceptual learning of specific item configurations also includes item-specific word form or not.

2.3. Perceptual Learning in Visual Word Recognition

The hypothesis that perceptual learning is also involved in visual word recognition has recently been proposed by Nazir and her colleagues (Nazir, 2000; Nazir et al., 1998). They introduced the perceptual frequency hypothesis of fixation location of visual word recognition: *"Word recognition] performance is best at the location where the eyes tend to land most frequently*" (Nazir et al., 1998, p. 820). Empirical evidence for this hypothesis stems from careful examination and modeling of the effect of fixation locations within words on word recognition performance.

This VPE in words consists of a systematic variation of word recognition performance as a function of fixation position within a word. When a skilled reader is asked to recognize a briefly presented five-letter word displayed such that the eye is fixating either the first, the second, the third letter of the word and so forth, a systematic variation of performance is observed: The word is recognized best when fixated slightly left of center and performance is better with the eyes fixating on the first compared to the last letter in the word. This pattern of results is reliable over a broad range of paradigms, such as the Naming Task (NT), the Lexical Decision Task (LDT), the Reicher and other Perceptual Identification Tasks (PITs), normal text reading and reading sentences for comprehension (e.g., Farid & Grainger, 1996; McConkie, Kerr, Reddix, Zola, & Jacobs, 1989; Montant, Nazir, & Poncet, 1998; Nazir, Heller, & Sussmann, 1992; Nazir et al., 1998; Nazir, O'Regan, & Jacobs, 1991; O'Regan & Jacobs, 1992; Underwood, Clews, & Everett, 1990) and has been shown for different dependent variables such as refixation probability, gaze duration, reaction times (RTs), and accuracy (Brysbaert, Vitu, & Schroyens, 1996; O'Regan & Jacobs, 1992; O'Regan, Levy-Schoen, Pynte, & Brugaillere, 1984; Vitu, O'Regan, & Mittau, 1990). Even children show the VPE from first grade on (Aghababian & Nazir, 2000; Nuerk, John, Jacobs, Schulte-Koerne, Graf, & Gauggel, 2000; Oßwald, Brustmann, Nazir, Nuerk, Graf, & Jacobs, 1998).

2.3.1. A Model to Account for the Viewing Position Effect

The observed asymmetry of the VPE for word recognition correlates with visual-field asymmetries in perceiving individual letters. The legibility of Roman letters – when embedded within other letters - drops faster with eccentricity when the stimulus is displayed in the left compared to the right visual field (Bouma & Legein, 1977; Nazir et al., 1992). A mathematical model that is based on this asymmetrical drop-off of individual letter legibility can describe the VPE sufficiently well in most cases (for a detailed discussion see Nazir et al., 1998), supporting the visual nature of the VPE in word recognition.

$$P_{d,A}(f,w,r_f) = \begin{cases} r_f \bullet \prod_{k=1}^{w-f}(r_f - kd) & : \quad f = 1 \\ \prod_{k=1}^{f-1}(r_f - kAd) \bullet r_f \bullet \prod_{k=1}^{w-f}(r_f - kd) & : \quad 1 < f < w \\ \prod_{k=1}^{f-1}(r_f - kAd) \bullet r_f & : \quad f = w \end{cases}$$

*Equation 1: The model has two free parameters and three function arguments. The three function arguments are fixed according to presentation conditions and two parameters are free to adjust the theoretical curve to actual data. The three fixed function arguments are fixation position f, word length w, and probability to recognize the letter at fixation r_f (which is set to 1) while the two free parameters are the drop-off rate to the right of fixation d and the asymmetry ratio A = drop-off rate to the left / drop-off rate to the right. k indexes the letter distance between the fixated letter and another letter to be recognized and possibly ranges from k = 1,...,w - 1. It is assumed that the probability to recognize a word on a given fixation position is computed as the product the independent probabilities of recognizing its constituent individual letters. The recognition probability of each letter is a function of its letter distance k from fixation. This probability drops from r_f = 1 with the drop-off rate d multiplied with k (see Table 1). E.g., if the drop-off rate d is 0.05, a letter that is k = 3 letter positions away from fixation is recognized with probability p = 1 – (3 * 0.05) = 0.85. However, empirically the drop-off rates to the left and the right are asymmetrical in letter recognition (Nazir et al., 1992). This asymmetry A is in letter recognition is responsible for the asymmetry of the VPE in word recognition (see Figure 1).*[1]

Fixation position in the word	Position of letter in the word					Probability of recognizing the entire word (product of the single probabilities)
	1	2	3	4	5	
1	**1**	.97	.94	.91	.88	.73
2	.946	**1**	.97	.94	.91	.78
3	.892	.946	**1**	.97	.94	.77
4	.838	.892	.946	**1**	.97	.69
5	.784	.838	.892	.946	**1**	.55

*Table 1. Example for the probabilities of recognizing a five-letter word (from Nazir et al., 1998) The probability of recognizing the directly fixated letter is set to 1. This probability drops linearly by d = .03 with each letter position of eccentricity to the right of fixation, and by A * d = 1.8 x .03 = .054 with each letter position of eccentricity to the left of fixation.*

[1] The model is unmodified from Nazir et al. (1998) but I have changed the labels of the parameters to obtain a more transparent correspondence to the interpretation and the name of the parameters. The drop-off rate (to the right) is now called d instead of b_r, the asymmetry between the drop-off rates is now directly introduced as A (instead of indirectly computed as b_l/b_r), the word length is called w instead of l and probability to recognition on fixation r_f instead of a. The numbering index is called k instead of n for contingency because later in this thesis the number of fixation points n_f will be introduced.

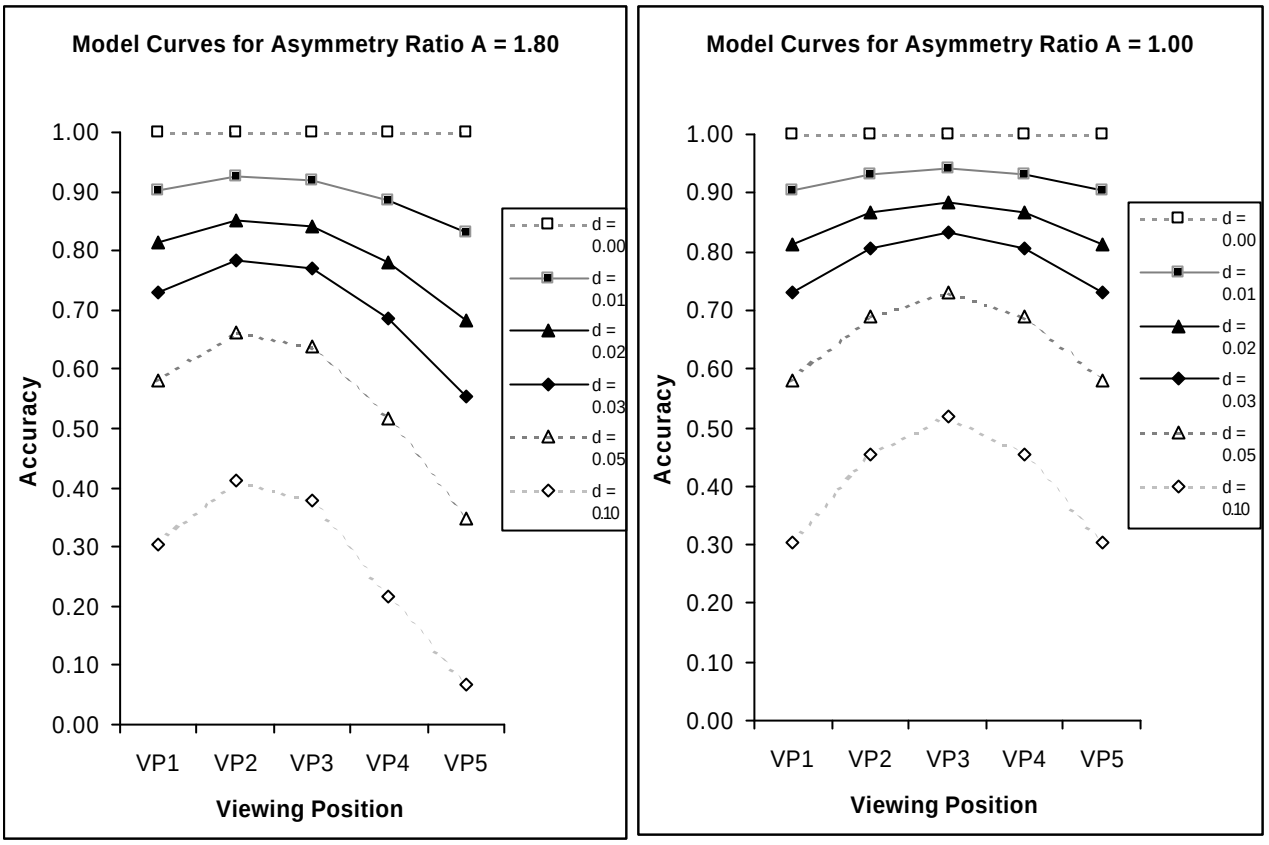


Figure 1: Model curves for different asymmetry ratios A and drop-off rates d for five-letter words. An asymmetry ratio of A = 1.80 (as in Table 1), coding asymmetrical drop-off rates of letter legibility, leads to asymmetrically shaped curves (left panel). However, if there is no asymmetry in letter recognition (A = 1.00), the word recognition curves remain completely symmetrical. The model is very constrained in that the drop-off rate d codes the height and the form of the curve. Higher drop-off rates may imply both, lower overall performance and greater difference between optimal and poor viewing positions. The possible problems of this double coding will be discussed when the model is used for derivation of hypotheses and when it is empirically tested.

2.3.2. The Perceptual Frequency Hypothesis of Location with Respect to the Model of Nazir et al. (1998)

The relevant issue for the validity of the perceptual frequency hypothesis whether asymmetry in letter and word perception fitted in the model by the asymmetry parameter evolves with reading habits (Nazir, 2000; Nazir, Deutsch, Grainger, & Frost, submitted; Nazir et al., 1998) or whether it can sufficiently be explained by other, item- and reading-habit-unspecific, factors, such as hemispheric differences in processing language (Banich, 1997; Brysbaert et al., 1996). Evidence for the perceptual frequency hypothesis of location comes from an effect that cannot be explained by hemispheric processing: Language-specific alteration of the VPE. The language-specific effect was found in a recent translingual study (Nazir, 2000; Nazir et al., submitted) with Roman and Hebrew scripts. Landing site distributions of the eye during reading are mirror reversed in these two scripts with opposite reading directions (left to

right vs. right to left; Deutsch & Rayner, 1999; Rayner, 1979). Perceptual learning effects, as a monotonic function of the frequency of retinal exposure, should therefore also show reversed patterns in the two languages. Indeed, performance was best for letter positions corresponding to the beginning of words in the respective script (Nazir et al., submitted).

Another consequence of the perceptual frequency hypothesis of location is that the visual asymmetry of the curve shifts as a function of word length. The most frequent landing site position of saccades in text reading is shifted more to the left in longer words (Nazir et al., 1998). For three-letter words the landing site distribution has a different profile. In contrast to longer words, the maximum of the landing site distribution of short words is shifted beyond the word center towards the end of the words (Rayner, 1979). If the perceptual frequency hypothesis is correct, this shift should be mirrored in the VPE. Thus, the asymmetry parameter in the viewing position model should change as a function of word length and language. The shift of asymmetry as a function of word length will be investigated in Study 2 in which word length is manipulated in addition to lexical frequency. This manipulation will to my knowledge represent only the second experimental test of the perceptual frequency hypothesis of location after the translingual study of Nazir and colleagues (Nazir et al., submitted). Fits with a generalized model and the analysis of the fitted asymmetry parameters will be a further tool to investigate the perceptual frequency hypothesis of location in Study 2. Having introduced the perceptual frequency hypothesis of *location* and evidence for it and its future investigation with experimental data and the viewing position model I now generalize the perceptual frequency hypothesis in visual word recognition to *item-specific word form.*

2.3.3. The Perceptual Frequency Hypothesis of Item-Specific Visual Word Form

Let me first note that the claim of my thesis is in contradiction with the literature, as most researchers today seem to agree that there is no item-specific visual word form processing, particularly no item-specific holistic processing, but that word recognition performance is solely determined by processing of abstract letter units (e.g., Besner & Johnston, 1989; Mayall et al., 1997; Paap et al., 1984). Investigation or discussion of visual word form processing has focussed on very different though not exclusive aspects: a) word shape or envelope shape (Haber & Haber, 1981; Haber, Haber, & Furlin, 1983; Paap et al., 1984; Wheeler, 1970), b) word-specific visual patterns or holistic word form (Allen, Wallace, & Weber, 1995; Besner, 1989; Besner & Johnston, 1989; Rudnicky & Kohlers, 1984), c) transletter or multiletter features (Besner & Johnston, 1989; Mayall et al., 1997), d) abstract single letter processing (Evett & Humphreys, 1981; McConkie & Zola, 1979; Paap et al., 1984; Rayner, McConkie, & Zola, 1980), e) the distinctiveness

of letter features and confusability or shape of specific letters (McClelland & Rumelhart, 1981; Paap, Newsome, McDonald, & Schvanefeld, 1982; Paap et al., 1984; Rumelhart & Siple, 1974; Ziegler, Rey, & Jacobs, 1998), f) lateral inhibition from neighbor letters (Mayall et al., 1997; Paap et al., 1984), and g) font-specific tuning or type typicality (Brown & Carr, 1993; Rudnicky & Kohlers, 1984; Sanocki, 1987, 1988). Hence, I need to define what I mean by visual word form in this thesis before I elaborate the perceptual frequency hypothesis of visual word form. For the present thesis I define visual word form of an item *to comprise the features (cf. Rumelhart & Siple, 1974), salient feature combinations (e.g., letters and letter combinations), their configuration in space (e.g., spacing) and the Gestalt formed by their overall configuration.*

Clearly, a more precise definition of 'visual word form processing' depends on the question of which of the above visual aspects determine best or exclusively word recognition performance. Isolating one or more possibly relevant visual properties is a difficult enterprise, because those aspects are not exclusive at all. Rather, they are often confounded in a natural way as, e.g., word shape and letter shape (Paap et al., 1984, for a delineation of these two aspects). However, possibly there is no single visual property that determines the ease of visual processing. Already in 1984, Rudnicky and Kohlers pointed out: *"Reading goes forward in many ways at once rather than through an orderly sequence of operations" (p. 231).* Mayall et al. (1997) reached a similar conclusion stating that *„lexical access is based on units coded at multiple levels."* Since recent research suggested that perceptual learning can occur at various levels in the visual system (Ahissar & Hochstein, 1997), a multiple levels approach would also seem most appropriate for a perceptual frequency hypothesis of item-specific visual word form.

While the multiple-levels approach of a perceptual frequency hypothesis is in line with most other accounts of visual word form processing, the postulate that perceptual learning is *also* item-specific is more controversial (e.g., Allen et al., 1995; Besner, 1989; Besner & Johnston, 1989; Haber & Haber, 1981; Haber et al., 1983; Mayall et al., 1997; Paap et al., 1984; Wheeler, 1970). Only two (a and b) of the seven upper aspects of visual word form processing are item-specific. The other five aspects (c – g) are general in the sense that they are not item-specific: Destruction of transletter features, abstract letter encoding, letter discriminability, the distinctiveness of letters, and case or font-specific tuning should affect the processing of all letter strings. Although different letter strings may be differentially affected, for example, by manipulations of letter discriminability (simply because they have different letters) those aspects imply no perceptual influence that is specific to the recognition of one particular word. Does the reading process also involve only those general visual aspects (i.e., c –g) or do we additionally use

information about item-specific visual word form? As Besner and Johnston (1989; see also Paap et al., 1984) pointed out, word-specific visual patterns are neither necessary nor sufficient for visual word recognition. They are not necessary, because case-mixed words can still be read rather easily. They are obviously also neither sufficient for the reading process in general, because nonwords can be read, nor are they specifically sufficient to account for case-mixing effects, because case-mixing affects nonword processing, too. Thus, reading can in principle be performed without computing item-specific visual word form. The critical question is, is this really the case?

As most researchers agree that there exists no item-specific visual word form processing I need to have a closer look at the form of reasoning used to reach this conclusion in order to argue why I still saw the possibility of finding item-specific word form effects and carried out this study.

2.4. Critical Evaluation of the Objections against Perceptual Frequency

2.4.1. Evaluation of Statistical Models and Theoretical Assumptions

The question whether the reading process is performed with or without computing item-specific visual word form is typically examined via the following form of indirect reasoning: General visual aspects (e.g., c –g) are manipulated in the same way (e.g., case mixing) for different stimulus groups (e.g., words of different frequencies and nonwords). Essentially, on the basis of a null interaction between visual manipulations and frequency it is concluded along AFL that item-specific visual form (i.e., one whole word feature) does not play any role. More specifically, the argument for a word vs. nonword manipulation runs as follows (Paap et al., 1984, p. 414). *„[..] Only words have word-shape feature. Because most nonwords are not represented in the lexicon, they do not possess shape features and, accordingly, nonwords can't be automatically activated by the perceived shape of a stimulus. This is a useful distinction because it suggests a critical test for any purported demonstration of word shape, namely, that the manipulation should affect words more than nonwords.“*

This critical test is usually statistically examined indirectly as an interaction in an analysis of variance (ANOVA). However, the derivation of test statistics and significance tests in an ANOVA relies on a particular theoretical model, the general linear model, which assumes that data can be adequately described as a linear additive combination of independent factors and their interactions. The interpretation follows, for the most part, additive factors logic: Additivity of effects of experimental manipulations on mean RT is taken to suggest that the underlying mechanism can be divided into independent operations (Roberts & Sternberg, 1993; Sternberg, 1969; see also Borowsky & Besner, 1993). However, other

models as, for example, the successful family of interactive activation models of word recognition (Coltheart & Rastle, 1994; Grainger & Jacobs, 1993, 1994, 1996; Jacobs & Grainger, 1991, 1992; Jacobs, Grainger, Rey, & Ziegler, 1998; McClelland & Rumelhart, 1981; Plaut, McClelland, Seidenberg, & Patterson, 1996; Ziegler et al., 1998; Zorzi, Houghton, & Butterworth, 1998) have given up a linear additive logic of independent factors. Instead, all sorts of information, visual, orthographic, phonological, and lexical can interact in a complex, dynamic way, but can also produce additive effects (e.g., Rumelhart & McClelland, 1982, Experiment 7).

This converges with Sternberg's early personal warnings when he introduced additive factors logic in 1969 arguing already that additivity on a given scale does not necessarily imply independence of factors or stages. That radically different mechanisms are capable of producing additive effects undermines the deterministic interpretation of a null interaction effect in an ANOVA as evidence against word-specific visual form. It must be considered that any null interaction between visual properties and frequency/lexicality obtained in an ANOVA can have two possible causes: First, that perceptual frequency does not play a role in visual recognition and, second, that the interpretation of the null interaction within the underlying statistical model, i.e. general linear model, and AFL may be inappropriate. Although the general linear model is – as the base model for the ANOVA – probably the model implicitly most often assumed to be true, this is not necessarily the case. In word recognition , in particular, alternative models such as interactive activation models have provided successful accounts of the results. However, these models make fundamentally different assumptions than the general linear model, which supposes that a given data pattern can be statistically tested as a simple additive equation of main factor effects and their interaction plus some error term. Given the importance of this, a more detailed elaboration of the theoretical consequences of an alternative interactive view, specifically for the item-specific word form effect, seems necessary.

Suppose, word recognition functions analogous to an interactive activation model framework and, both, item-specific perceptual and lexical frequency, influence word recognition performance in a monotonic way: The higher either frequency the better word recognition performance will be. Lexical frequency operates both via higher resting levels and top-down feedback, and can be used to compensate for less salient or incomplete visual information. Evidence for this view comes from studies of the VPE: The VPE is much less pronounced in models and empirical data when there is top-down processing using lexical frequency information than when there is no top-down processing (Montant et al., 1998). On the least optimal viewing positions on which visual information is most difficult to integrate, top-

down feedback particularly helps to compensate for the poorer visual conditions. Thus, if visual information is distorted at the perceptual level, e.g., by case-mixing, top-down feedback can help to compensate for that distortion (Ziegler et al., 1998). As words with higher frequency are processed by units with higher resting activation levels (Jacobs et al., 1998; McClelland & Rumelhart, 1981) top-down feedback is especially powerful for them and strongly compensates for distorted visual information.

Item-specific perceptual frequency of visual word form, on the other hand, is not existent for nonwords (see Paap's argument above) and should have a particularly powerful effect for high-frequency words according to the perceptual frequency hypothesis. Hence, in the standard visual word form study, one may have to deal with two opposite interacting processes: A distortion of perceptual shape that should affect performance most for high-frequent words and least for nonwords, and, lexical top-down processes that might compensate for perceptual distortion and should be most powerful for high-frequent words and least powerful for nonwords.

Now, the question arises which of the two opposite processes is stronger: If top-down feedback compensation is stronger, one should observe underadditivity, i.e., high-frequent words should be less affected by visual distortion than low-frequency words and nonwords. If the influence of item-specific perceptual distortion is stronger, then one should obtain overadditivity with high-frequent words being more affected by distortion than low-frequent words and nonwords. Finally, if the two opposite processes cancel each other out, one should observe an additivity pattern. Users of interactive activation models know that answers to this question based on computational evidence depend on the actual implementation of perceptual distortion (for different implementations see Mayall & Humphreys, 1996b; Montant et al., 1998; Ziegler et al., 1998), the parameter setting, and the implementation of task-specific processes and the strategic demands (Grainger & Jacobs, 1996; Jacobs & Grainger, 1994, for a discussion of these issues).

In sum, those considerations show that a null interaction in an ANOVA based on the general linear model is not necessarily conclusive. In particular, in domains where alternative models other than the general linear model have provided successful accounts for the data, its interpretation is problematic, since the general linear model may no longer be appropriate. Nevertheless, indirect reasoning from a null effect of an ANOVA interaction could provide some constraining results, if both, the empirical evidence and the model predictions, were unambiguous. Unfortunately, they are not.

2.4.2. Evaluation of the Empirical Evidence

Similarly to computational studies, the results of experimental investigations of the interactivity of frequency and stimulus distortion depend on the type of distortion, task and strategic demands of the task (for reviews, see Allen et al., 1995; Besner & Johnston, 1989; Mayall & Humphreys, 1996a). The interaction between lexical frequency[2], hereafter just called frequency as in the literature, and case-mixing differs between tasks (Besner & Johnston, 1989; Besner & McCann, 1987; Mayall & Humphreys, 1996a), but more critical for my argument is that even within one task, the results are inconsistent. Mostly, approximately additive effects of case-mixing and frequency are reported in the LDT (Besner & McCann, 1987; Frederiksen, 1978; Mayall et al., 1997; see also Besner, 1989, for function words). However, sometimes significant interactions between case-mixing and word frequency are found (Kinoshita, 1987; Allen et al., 1995; and a trend in Experiment 4 of Mayall & Humphreys, 1996a). While Kinoshita (1987) reports that nonwords are more affected than words (see also Mayall & Humphreys, 1996a, Exp. 4), Allen et al. (1995) obtained just the opposite result for short presentation times.

Of particular interest are the results of Kinoshita (1987). Kinoshita pointed out that his nonwords were much more similar to words (many were word neighbors) than Besner and McCann's (1987) and discussed the diverging results with respect to different decision strategies. This converges nicely with simulations in interactive activation models. For example, the Multiple Read-Out Model (MROM; Grainger and Jacobs, 1996) can not only account for the well-known better performance in the LDT with easy nonwords (e.g., Balota & Chumbley, 1984), but also for interactions of nonword lexicality with other linguistic effects. In particular, the neighborhood density effect for words could only be obtained with word-unlike nonwords, but not with word-like nonwords (Grainger & Jacobs, 1996; Experiments 1b and 1c). The model captures this effect by lowering an intralexical familiarity criterion for experiments in which word-unlike nonwords are presented (see also Balota & Chumbley, 1984; Kinoshita, 1987). Thus, interactive activation models can, in principle, account for diverging results within the same task depending on nonword list properties. Therefore, it is worthwhile considering interactive models as possible alternative accounts to current additive factor models of the diverging effects in case mixing studies, since additive factor models cannot deal with such divergence and can only account for the most frequently found additivity in such studies.

---

[2] With the term frequency manipulation I want to denote both the manipulation of frequency for words (high- vs. low-frequency words) and the manipulation of lexicality (word vs. nonword as a stimulus with frequency 0.00) in the subsequent paragraph. This is done because with regard to the current issue these manipulations are interpreted in the same way and the same logic holds for both manipulations.

2.4.3. Summary and Conclusions

In the above paragraphs, I meant to raise some doubts about the validity of the rejection of the item-specific word form hypothesis, based on AFL: Top-down processes of lexical frequency in interactive activation frameworks could well mask effects of item-specific distortion of perceptual frequency as both processes tend to work in opposite directions and interact in a complex way. A rejection based on AFL is further questioned by diverging results even within one task and not only between tasks. The fact that interactive activation models have been shown to account for different results within one task in other areas of visual word recognition supports the hypothesis that we look at an interactive process. This interactive process of perceptual frequency and lexical frequency cannot easily be disentangled by indirectly investigating an interaction in the general linear model using AFL. Instead, it would be more advantageous to directly investigate perceptual frequency with a perceptual frequency manipulation that is not confounded with lexical frequency. This manipulation should not be detrimental to all words and nonwords and just be implementable as perceptual noise in a model (e.g., Mayall & Humphreys 1996b). Rather, it should have a specific influence on specific groups of words with a specific perceptual form. The German language provides such a possibility, that I will use for the investigation of visual word form in Experiment 1.

The predictions of the model of Nazir et al. (1998) are straightforward. Because no metric and no parameter for item-specific perceptual form is incorporated, the results between different word forms should not differ. Additional to this null effect prediction, the model can be very helpful, as alternative letter legibility accounts of my experimental manipulation can be tested with the model. However, how the model can help to dismiss alternative accounts can only be elaborated when the specific word form manipulation I use for German words is introduced. This will be done in the introduction directly prior to Experiment 1.

2.5. The Confusability Effect in Visual Word Recognition

I have introduced two visual effects above, the well established VPE, and the neglected item-specific word form effect (e.g. Besner, 1989; Kinoshita, 1987). However, when I examine the alteration of the lexical frequency effect and the neighborhood effect by visual variables, I will additionally use a third visual effect for this examination. In contrast to the VPE, letter legibility is not manipulated in the same way for any letter. Rather, *the confusability effect refers to the specific attributes of a given letter: The ease with which it can be confused with other letters* (also sometimes called similarity effect). Letter

confusability has been shown to influence visual word recognition, both, empirically and in modeling (for its investigation see, e.g., Arguin & Bub, 1995; Jacobs & Grainger, 1991; Paap et al., 1984; Schmidt-Weigand, Rey, Nuerk, Graf, Jacobs, & Van Orden, 1998; Ziegler et al. 1998; for models see Bouwhuis & Bouma, 1979; Grainger & Jacobs, 1996; McClelland & Rumelhart, 1981; Paap et al., 1982; Rumelhart & Siple, 1974; Ziegler et al., 1998). Very early, letter confusability has been demonstrated to potentially drive other effects. In a carefully controlled series of experiments Paap et al. (1984) investigated the question whether previously reported effects of whole word shape (Smith, 1969; Haber et al., 1983; see above paragraphs) hold if other factors are controlled. In a proofreading task, Paap and colleagues found that a suspected word shape effect was rather due to letter confusability (or similarity, Experiment 1). Other authors investigated the effects of letter confusability in priming paradigms. Jacobs and Grainger (1991) found stable priming effects of visually similar letters in an alphabetical decision task, in which participants were required to classify a character either as a letter or a non-letter. If the primed letter shared many features with the probe letter, latency was faster for the probe letter than if it did not. However, in a letter identification task, the effect is reversed. There, physically similar primes inhibit naming times (Arguin & Bub, 1995). Finally, in two PITs, letter confusability was the best regression predictor of performance (Schmidt-Weigand, 1999; Schmidt-Weigand et al., 1998; Ziegler et al., 1998).

That letter confusability could play a role in visual word recognition is consistent with the architecture of some (e.g., Grainger & Jacobs; 1996; Jacobs et al., 1998; McClelland & Rumelhart; 1981; Paap et al., 1982) but not all models of visual word recognition (e.g., Plaut et al., 1996). If letter confusability has an impact on frequency and neighborhood effects, the generality of any model without an architecture for confusability is severely limited to one certain presentation condition. Particularly, the generality of simulations of frequency and neighborhood effects is questionable if letter confusability interacts with other standard manipulations in a non-additive, non-linear fashion. Null effects may then not be due to the specific non-visual manipulation used in the study but rather to the particular letter confusability or legibility interacting with this manipulation. In the next three chapters, I will present evidence that even the standard hallmark effects of visual word recognition, the frequency effect, the word length effect, and neighborhood effects are sensitive to the visual conditions under which they are investigated. All these effects and most of their interactions seem to diminish, when they are investigated under near-optimal visual conditions. Holding visual presentation conditions constant at a high legibility level as it is done in most laboratory experiments may produce null effects that are specific to those particular visual

conditions. If even these standard effects can change as a function of presentation conditions, the importance of visual determinants would become quite obvious.

3. On the Alteration of Lexical Frequency Effects by Visual Determinants

3.1. Introduction and Definition

The lexical[3] frequency effect (hereafter frequency effect as it is called in the literature) is probably the most reliable and most prominent effect in visual word recognition. The frequency of a word refers to the number of occurrences in written and/or spoken language (or any combination dependent on the database used) and is usually indexed per million occurrences. The frequency effect denotes better word recognition performance for more frequent words and is found reliably in virtually all word recognition tasks. In contrast to the visual perceptual frequency effect, it is here not the frequency effect itself that is controversial but the nature of its interaction with other variables.

In this review, the idea that the frequency effect and its interactions with other variables can be generally studied using one certain isolated visual condition will be questioned. Instead, evidence will be presented that the frequency effect and its interactions become larger as presentation and stimulus conditions become poorer. Hence, failures to observe the frequency effect itself or its interactions with other variables are, according to this hypothesis, due to optimal viewing, presentation or skill conditions, that do not necessarily occur in normal reading. That the frequency effect itself and its interaction with other variables can rather be found under non-optimal viewing conditions may pose a problem for models that assume sequential, non-interacting stages of visual and lexical processing (Besner & Smith, 1992; Borowsky & Besner, 1993). However, it is compatible with he architecture of interactive models that assume interactive activation exchange between visual and lexical levels (e.g., Jacobs et al., 1998). Sequential stage models and interactive accounts would make different predictions with regard to the viewing position effect. At the end of the chapter, I will examine how the parameters of the model of Nazir et al. (1998) would change in different experimental conditions if either of these two accounts would be

[3] The common frequency effect will be termed lexical frequency effect whenever it shall be distinguished from perceptual frequency effects in this thesis. Note, however, that some authors (e.g., Paap et al., 1982; Paap & Johansen, 1994) argue that the common frequency effect is rather post-lexical than lexical in nature. However, because most authors understand the frequency effect as a lexical effect, the attribute lexical rather than post-lexical was chosen to distinguish lexical from perceptual frequency throughout this thesis.

true. The way in which the parameters of the model do or do not change will provide constraints for the model refinement at the end of this thesis.

3.2. Interactions and Confounds of Frequency Effects with Other Stimulus Attributes, Task Demands, and Reading Skill

The word frequency effect is probably the most prominent effect in the visual word recognition literature and, in contrast to neighborhood effects, also a very reliable one. Word frequency affects RT or accuracy performance in virtually all standard visual word recognition tasks like the NT and the LDT (e.g., Balota & Chumbley, 1984; Balota & Spieler, 1999; Forster & Chambers, 1973; Hudson & Bergman, 1985; Lee, 1999; Weekes, 1997), different kinds of PITs (e.g., Nazir et al., 1998; Rey, Chesnet, Bijeljac-Babic, Grainger, & Jacobs, submitted; Ziegler et al., 1998). Frequency effects were not only found with standard dependent variables like RT or accuracy but also with all kinds of oculomotor variables like gaze duration, fixation duration, or refixation probability (e.g., Hyönä, 1995; Inhoff & Rayner, 1986; McConkie et al, 1989; Vitu, O'Regan, Inhoff, & Topolski, 1995; Vitu et al., 1990) as well as with electrophysiological variables such as the latency of early negativity (Osterhout, Bersick, & McKinnon, 1997).

From the first studies examining the word frequency effect, this effect has been shown to be both, task-overlapping and task-sensitive. Forster and Chambers (1973) found correlations between responses to identical items in the NT and the LDT of r = .55 and claimed that those correlations were due to differences in word frequency (but see Hudson & Bergman, 1985, for a thorough investigation leading to different results). However, when the sizes of frequency effects have been compared between these two tasks usually a bigger effect of frequency has been found in the LDT (Balota & Chumbley, 1984; Frederiksen & Kroll, 1976; see O'Regan & Jacobs, 1992, for a short review). This bigger frequency effect in the LDT has led researchers to propose an extra familiarity assessment mechanism in the LDT, not included in the NT (Balota & Chumbley, 1984; see Grainger & Jacobs, 1996, for a computational implementation of a similar mechanism). Recently, regression studies have tried to go "beyond measures of central tendencies" (Balota & Spieler, 1999, p. 32). They have demonstrated that word frequency can affect different distributional parameters in an ex-gaussian distribution that are specific to the reading task. While in the NT frequency affected only the μ parameter (i.e., the mean of the normal distribution being part of the ex-gaussian convolution), it affected both, μ and τ (the parameter of the exponential distribution being part of the ex-gauss convolution) in the LDT. Balota and Spieler argued that μ is constant across tasks while τ differs between LDT and NT and related these distributional parameters to

Balota and Chumbley's earlier 2-stage model. In sum, the word frequency effect is found in virtually all word recognition tasks (see, e.g., Inhoff & Rayner, 1986; Paap & Newsome, 1980, for some exceptions; Paap & Johansen, 1994, for an interpretation along verification accounts, and Grainger & Jacobs, 1996; Jacobs & Grainger, 1992, for an accommodation of null effects in the Reicher task by an interactive framework). However, its size seems to differ greatly between tasks and across different dependent variables and, therefore, it is a viable source for computational and distributional modeling.

However, even a probable word frequency effect can sometimes be due to a confound with another stimulus attribute. Consider, for example the correlation between the NT and the LDT data that Forster and Chambers (1973) attributed to the frequency factor. Hudson and Bergman (1985) reanalyzed their data, performed own experiments, and found that partialling out the frequency did not significantly alter the correlations between LDT and NT latency found by Forster and Chambers. This suggests that frequency did not play a major role for that correlation. Using log word frequency for the partial correlation altered the result a little more, yet partial and regular correlation did still not differ significantly (Hudson & Bergman, 1985, p. 48). In their own experiments, they found that correlations of reaction time (RT) with frequency in the NT (but not in the LDT) were rather low, when length was partialled out. The opposite was not true: Partialling out frequency did not much alter the correlation with length in their data. Similarly, O'Regan et al. (1984) confounded length with frequency, making a replication of their NT experiment necessary (O'Regan & Jacobs, 1992). Thus, even the word frequency effect may sometimes only be found because of confounds with other measures. However, usually if an interaction of frequency with word length is observed the two effects interact in a characteristic way. The interaction is overadditive, i.e., low-frequency being most detrimental for long words, or vice versa, length particularly seems to affect low-frequency words. This interaction is reviewed in more detail below.

Besides task differences, the second constraining source for modeling visual word recognition are (null) interactions of frequency with other stimulus attributes. However, these interactions are often interpreted with the focus lying on the differences between other stimulus attributes within stimulus groups of the same frequency. The reversed focus on different frequency effects within other stimulus groups is less often taken. Apart from word length, orthographic-phonological properties of stimuli present an example for that line of reasoning. It is argued that consistency, regularity or subcomponent frequencies only or particularly affect low-frequency words (e.g., Jared, 1997; Jared, McRae, & Seidenberg, 1990; Nuerk, Rey, Graf, & Jacobs, 2000; Stone, Vanhoy, & Van Orden, 1997; Treiman, Mullenix, Bijeljac-Bibac, & Richmond-Welty, 1995; Ziegler, Montant, & Jacobs, 1997). Rarely, if ever, the

other direction of the interaction is interpreted, i.e., how frequency effects are altered in different stimulus groups, e.g., by consistent or inconsistent stimuli groups. However, this is the perspective, which I will adopt.

Using this perspective, the literature suggests that word recognition performance seems to be more affected by frequency manipulations in those participants whose overall skill and reading performance is poor. More than 20 years ago, Perfetti, Goldman and Hogaboam (1979) found that unskilled young readers (as measured in a reading test) showed greater frequency effects than skilled readers in a NT (Exp. 3). Moreover, unskilled readers were also more likely to show interactions of frequency with other variables such as word length and context in the expected direction. Unskilled readers suffer particularly from a combination of disadvantageous stimulus attributes. As concerns long words, unskilled readers were particularly inferior for low-frequency items in all three experiments. In a similar way as in this interaction with word length and skill, frequency interacted with context and skill. If three-way interactions were found, frequency was particularly important when all other conditions were bad, i.e., in the no context (isolated reading) condition for unskilled readers. This result found for young readers is congruent with findings of adult readers, for which frequency and visual quality only interacted in participants who were more error-prone (Plourde & Besner, 1997; see below for details). In contrast, an older study failed to find an interaction between frequency and skill for undergraduate students with skill being measured by a vocabulary test (Butler & Hains, 1979). However, the observation that frequency interactions are particularly likely in readers with bad reading skills is also supported by a recent study on pure alexia (Behrmann, Plaut, & Nelson, 1998). They found a systematic increase of frequency effects with word length in their patients.

Altogether, these results from beginning readers and patients with reading disorders indicate that non-optimal reading skills are more likely to either increase frequency effects themselves or to increase frequency effects under particularly non-optimal stimulus or presentation conditions. Furthermore, frequency effects tend to become smaller as stimulus attributes become poorer. As a corollary this implies: Null effects of frequency obtained under optimal stimulus conditions in optimally skilled readers may not generalize. As this thesis is concerned with the alteration of frequency by visual determinants, I now discuss whether there is evidence that this corollary extends from stimulus attributes and particpants' skill to visual presentation conditions.

3.3. Interactions of Frequency Effects with Visual Presentation Conditions

3.3.1. Interactions of Frequency with Visual Distortions

Interactions of frequency with visual distortions have in large part already been discussed in the chapter about the importance of visual determinants in visual word recognition. However, in the above chapter, the perspective was different: I discussed whether visual attributes of stimuli, in particular item-specific word form, is altered as a function of frequency or not. The perspective of this section, however, is reversed. I discuss how the frequency effect is altered as a function of visual determinants.

If my above hypothesis is also true for the interaction of frequency with visual determinants, any interaction should be in the following direction: The frequency effect should be larger when visual presentation conditions make word recognition more difficult. However, as introduced above, sometimes null interactions of frequency with visual distortions have been observed. Again, there is evidence that these null interactions depend on the specific stimulus and presentation conditions. Moreover, as for perceptual frequency, there is evidence that failures to detect interactions of lexical frequency with other variables may be rather specific to presentation conditions and analysis strategies used in common laboratory experiments. I will argue that these experiments are not conclusively indicating a general non-existence of these interactions.

As reviewed above, there are a lot of studies reporting null interactions of frequency with visual determinants. Very early, a dissociation was claimed between a frequency by context vs. a frequency by stimulus quality interaction. Becker and Killion (1977) found that frequency did not interact with visual stimulus quality while the latter did interact with context. They concluded that their results supported models that assume a non-interaction between frequency and visual stimulus intensity (see also Borowsky & Besner, 1993). However, they seem to have assumed compensating influences of expectancy of a stimulus in a similar way an interactive model would assume such influences for frequency*: "It seems that the processing of an expected stimulus can somehow compensate for the otherwise disruptive effects of poor stimulus quality" (p. 400)*. Vice versa, interactions between frequency and context in the absence of interactions between frequency and stimulus quality have also been reported (Becker, 1979; Borowsky & Besner, 1993). Borowsky and Besner (1993; see also Besner & Smith, 1992) interpreted these findings using the AFL and concluded that context and visual quality affected a common stage of processing while frequency and visual quality did not. Moreover, another

form of stimulus degradation, case mixing, often revealed additive effects of case mixing and frequency or lexicality (Besner, 1989; Besner & McCann, 1987; Frederiksen, 1978; Mayall & Humphreys, 1996b).

Thus, on the basis of these studies I would not expect an interaction of frequency with visual quality if two aspects did not make the above results controversial. First, in all of these paradigms, diverging results have been found dependent on stimulus and presentation condition and participant groups. Second, as already discussed with regard to perceptual frequency, finding a null interaction in the ANOVA is not necessarily conclusive. If the AFL (Sternberg, 1969) is used to show that additive sequential stages are involved, it is implicitly assumed that the general linear model is true and that AFL can be applied. These two points will be elaborated below.

For both, visual quality manipulations and case mixing manipulations, different results have been obtained. For example, Norris (1984) found interactions of stimulus quality and frequency in a LDT (see Wilding, 1988, Experiment 2, for similar results). In case mixing studies frequency and/or lexicality have also been shown to interact with case mixing under some circumstances. (Allen et al., 1995; Kinoshita, 1987; and a trend in Experiment 4 of Mayall & Humphreys, 1996b). While Kinoshita (1987) reports that nonwords are more affected by case mixing than words (see also Mayall & Humphreys, 1996b, Exp. 4), Allen et al. (1995) reported just the opposite for short presentation times. In addition to case mixing, presentation conditions also seem to affect frequency effects. For example, Borowsky and Besner (1993) point out that both Norris (1984) and Wilding (1988) used long warning-signal-target intervals and found interactions of frequency with visual determinants, while in other settings such an interaction was not found (see section above). Moreover, when other visual manipulations were used such as the manipulation of viewing position in isolated word reading, frequency effects seemed to be stronger when the viewing position was non-optimal (e.g., O'Regan & Jacobs, 1992; see introduction to Experiment 1, in which viewing position manipulation is used, for details). Thus, interactions of visual determinants with lexical frequency seem to depend on specific stimulus and presentation characteristics not only in the special case of perceptual frequency but also more generally.

In the above cited studies the question whether two variables affect the same stage or process is usually investigated and statistically examined indirectly as an interaction in an ANOVA. However, the ANOVA relies on a statistical and theoretical model, the general linear model, which assumes that data can be adequately described as a linear additive combination of independent factors and their interactions: Additivity of effects of experimental manipulations on mean RT is taken to suggest that the underlying processing mechanisms can be divided into independent operations (Roberts & Sternberg,

1993; Sternberg, 1969), an assumption that is not shared by interactive models (McClelland & Rumelhart, 1981; Plaut et al., 1996; Jacobs et al., 1998). These models can also produce additive effects (see, e.g., Rumelhart & McClelland, 1982, Exp. 7, and introduction above).

In the last years, researchers have become more aware that comparing measures of central tendencies in an ANOVA may not be sufficient to detect word recognition effects and began to examine different distributions and other parameters than the mean of the distribution (Balota & Spieler, 1999; Plourde & Besner, 1997; Schmidt-Weigand et al., 1998). When reviewing task differences, I had introduced Balota & Spieler's (1999) finding that the difference between the frequency effects in the NT and the LDT is not appropriately described in terms of changes in measures of central tendency. With regard to this paragraph, their study also suggests that applying the general linear model to measures of central tendency only may not be sufficient to discover interactions between frequency and other factors.

Even more interesting are the findings of Plourde and Besner (1997) in which they consider their own earlier interpretation of null interactions using the general linear model as at least not being uncontroversial. *"Another possibility is that it is premature to reject the activation models' account of word frequency based on the finding that word frequency and stimulus quality are additive factors"* (p. 183). When they fitted the parameters of the ex-gaussian distribution[4] to their LDT RT data, they found ambiguous results. Frequency and stimulus quality both affected the μ and τ parameter of the ex-gaussian distribution. Thus, both effects are not restricted to a simple shift in distribution (as measured and tested in the ANOVA) but also produce longer tails and more asymmetric distributions. The shortcomings of a conventional ANOVA are nicely demonstrated by their study because in the overall ANOVA two simple main effects (i.e., a pure mean shift in the RT distributions) were found and nothing else (i.e., no indication of a change in the τ parameters). Even more problematic for simple linear models are the error analyses: They found an interaction between frequency and stimulus quality in the expected direction. Low-frequency and highly degraded stimuli yielded particularly high error rates. Finally, they divided their participants into a low and a high error group. In sum, the high error group showed interactions between stimulus quality and frequency in an overall ANOVA over mean RT and accuracy, while the low error group did not show any interaction in any dependent variable. The analysis of the ex-

[4] Note that the ex-gaussian distribution function may still not be the true underlying distribution function for the data. Even if one assumes that a true underlying distribution function exists, it may be impossible to find out which distribution function is the true one because it is not always the one that fits best (see the model section of the general discussion for elaboration). However, the analyses with the ex-gauss function nicely demonstrate that the normal distribution function underlying the ANOVA with the mean as the only examined parameter may not be true distribution function, and not even the most appropriate one to detect differences between conditions.

gaussian RT distributions revealed one important difference between the two error groups – stimulus quality significantly affected τ and the asymmetry of the distribution in the high error group but not in the low error group. This might be a possible source of the different interaction patterns in the two groups. These analyses support the idea that the detection of an interaction between frequency and stimulus quality depends on the participants, dependent variables and analyses used in an investigation.

However, even those distribution analyses may not be appropriate. Behrmann et al. (1998, p. 42) argue with respect to reading deficits in pure alectics. *"There may be an inverted U-function relating the severity of impairment to the strength of higher level effects in LBL readers. When the visual input is well processed, top-down support is largely unnecessary, and when the deficit is too severe, higher-order representations are not strongly engaged."* An analogous argument may hold for difficult stimulus and presentation conditions. At least when accuracy is the dependent variable this is trivial. Although frequency effects may increase with poorer presentation condition, naturally, no frequency effects will be observed any more, when presentation conditions are too difficult, i.e., when the stimulus can barely be identified. Clearly, if such an argument would be true for other dependent variables than accuracy, frequency interactions with visual variables would only be observed in a middle range, in which stimulus attributes, presentation conditions and participants' skills are neither optimal, nor too poor. A failure to find interactions of frequency with visual variables would then only account for the actual stimulus, presentation and participant skill combination in the very experiment. However, it would hardly allow general statements about the role of lexical processes independent of this combination.

3.3.2. Interactions of Frequency with Viewing Position

Because viewing position is a visual variable that is manipulated in all three studies of this thesis, the interaction of frequency with viewing position is reviewed in more detail. If frequency effects generally become greater as visual presentation conditions become poorer, then frequency effects should be greatest on poor viewing positions. Indeed, the review will show that if any interaction of viewing position with lexical variables is found within an ANOVA, then it is in the expected direction: Lexical information generally seems to help more, when viewing position conditions are poor.

The simplest case of a viewing position manipulation is to present words in one or the other hemifield or hemispace. Unilateral presentations of linguistic stimuli usually yield a right hemifield advantage in visual word recognition which commonly is assumed to reflect a left-hemisphere superiority in right-handers (for reviews, e.g., Banich, 1997; Bradshaw & Nettleton, 1983; Springer & Deutsch, 1998). As

concerns frequency, the results are inconsistent and seem to depend on the task involved. While Iacoboni and Zaidel (1996) failed to find a hemifield by frequency interaction in the LDT, Rastatter and Loren (1988) reported such an interaction in the NT. The interaction was in the expected direction: Inferior viewing positions (i.e., left hemifield) led to greater frequency effects. In a meta-analysis combining the sample of Iacobini and Zaidel (1996) and their own, Weekes, Capetillo-Cunliffe, Rayman, Iacoboni, and Zaidel (1999), however, also found such an interaction between visual field and frequency, although it was not present in all of their new experiments either. Thus, overall, frequency effects are larger when words are presented in the inferior visual field.

When viewing position is more continuously manipulated, interactions of viewing position with frequency are in that same direction. O'Regan and Jacobs (1992) found interactions of frequency with viewing position in both, the LDT and the NT, in the expected direction. The frequency effect in the LDT was stronger in the (inferior) end position than in the middle position that is closer to the optimal viewing position. If a significant interaction between frequency and viewing position for a single word length was found in the NT, then it was again in the expected direction – the smallest frequency effect was obtained on the middle viewing position. Similar interactions were found for refixation probabilities (McConkie, et al., 1989). In their mathematical model predicting refixation probabilities, frequency did not only affect the overall likelihood of refixations but – when word length was held constant to avoid possible confounds – it did particularly affect the likelihood of refixations away from the word center (see p. 249). Again, frequency had particularly large effects on inferior viewing positions. Related effects were also found by Pynte (1996). He varied the information distribution within a word. Fixations on inferior viewing positions within a word (i.e., at letter positions not dissolving the ambiguity between the target and lexical competitors) affected high- and low-frequency words differently. Again, inferior fixation positions lead to more refixations, especially in low-frequency words.

However, in continuous text reading the interaction between lexical and visual factors is more controversial (e.g., Vonk, Radach, & van Rijn, 2000, for a review) and the results seem to vary for different dependent variables. According to the processing difficulty hypothesis (Hyönä, 1995; Hyönä & Pollatsek, 1998) a parafoveally presented low-frequency word will narrow down the span of effective preprocessing which in turn leads to a shorter forward saccade and a different landing position. For example, Inhoff and Rayner (1986) found stronger frequency effects on gaze duration and first fixation duration and different fixation duration distributions when parafoveal preview was possible compared to a condition were it was not possible. This implies that frequency again particularly exerts its effects in poor

visual conditions (i.e., in the parafoveal preview condition) but not in a control condition (without parafoveal preview). However, the results in this domain are ambiguous (see Rayner, Sereno, & Raney, 1996; Vonk et al., 2000, for a review of diverging results). The reason may be that in continuous text reading the matter is complicated by the problem of how many words are processed and preprocessed and how that number influences word recognition attributes (Deubel, O'Regan, & Radach, 2000).

In sum, the discussed results seem to imply that frequency affects word recognition more when viewing position is inferior. This is in line with connectionist model simulations (Montant et al., 1998), in which the word-letter feedback parameter was manipulated. However, again, the results depend on the specific implementation of presentation and stimulus conditions and the dependent variable used. In continuous word reading the effects of frequency, information distribution and landing site distribution of both, the parafoveally previewed and the previously processed word seem to influence word recognition in a way that is not completely resolved yet. Thus, one is currently on safer ground when the interaction of frequency effects and viewing position is examined in isolated word reading than in continuous word reading. Therefore, I restrict my investigation of the nature of the interaction between frequency and viewing position effects to isolated word reading conditions.

3.4. Objectives and Hypotheses Derived from the Model of Nazir et al. (1998)

Generally, in reading research the assumption is made that if a word is presented at central fixation in one of the standard word recognition tasks, visual factors can be neglected except in masked priming tasks (e.g., Forster & Davis, 1991; Jacobs & Grainger, 1991). However, there is some evidence that visual determinants such as viewing position interact with lexical attributes of a word. Therefore, the generalization of a lexical effect or null effect found on central fixation position to any (more inferior) viewing position may be premature because frequency effects become stronger as viewing position becomes poorer. However, I want to investigate this interaction not only ordinally but also quantitatively. This may lead to more specific hypotheses about the variation of the free parameters in different conditions, the expected goodness of fit of the model of Nazir et al. (1998), and the constraints for its future refinement.

The drop-off rate should decrease with increasing frequency because the drop-off rate is the only free parameter that currently can influence the height of the curve. However, the specific incorporation of the drop-off rate in the model predicts a change not only in height but also in the form of the curve. The curve should become flatter as stimulus frequency increases. So, if the empirical curves for high and low-

frequency words would not differ in form (e.g., if visual and lexical parameters are just additive factors), the model - by its very architecture - would fail to fit these data. In the refined model, a lexical parameter shall be added, so that variation in lexical factors do not affect visual parameters such as the drop-off rate. The shape of the viewing position curves for different frequencies will thus provide clear constraints for this incorporation. If the shapes of the curves do not differ between frequency groups a lexical additive parameter will be sufficient. In contrast, if the shapes do differ, a purely additive lexical parameter is not sufficient. On the basis of the reviewed literature, I expect that the lexical parameter will not be simply additive in that it increases height for all viewing positions equally.

If the interaction between viewing position and frequency is as expected, my hypothesis must be validated in a second task to allow for more general task-overlapping conclusions. In a word fragmentation task visual letter confusability and frequency have been shown to be strongest predictors for performance in a regression analysis (Ziegler et al., 1998). Therefore, I chose this task for a cross-validation of the hypothesis that frequency helps most, when visual presentation conditions are poor.

Finally, the interactions of frequency with other variables such as word length should be particularly high on poor viewing positions. To see if this is the case I will manipulate word length in addition to frequency and viewing position to investigate the interaction of word length and frequency on different viewing positions. Therfore, I will give a review of the literature about word length effects, about their interactions with other variables and with frequency, in particular, and about how these interactions vary with visual presentation conditions.

4. On the Alteration of the Word Length Effect by Visual Determinants

4.1. Introduction

In the last chapter, I suggested that not only frequency effects themselves but also interaction effects of frequency with other variables may increase under poor viewing conditions. The interaction which will be considered in this thesis is that of frequency and word length. I will try to show that word length effects tend to increase when stimulus conditions, presentation conditions and participants' skill are poor. In particular, they tend to increase with low frequency and with poor viewing position. In one study, the interaction between word length and frequency was particularly observable in bad viewing positions. Because the model of Nazir et al. (1998) can yet only fit data for which the number of viewing positions equals word length, it was necessary to generalize this model to fit data from any combination of the number of viewing positions and the number of letters in one experiment. This will be done in this chapter.

Based on that generalization the hypotheses concerning the parameter values of the model fits and their goodness of fit for the word length effect and its interaction with frequency will be specified.

4.2. Interactions with Other Variables

Similarly to frequency effects I suggest that word length effects are more likely to occur when stimulus attributes, presentation conditions and participants' skills are non-optimal. Word length effects have been found in virtually every task and for each dependent variable of visual word recognition such as RT in the NT and the LDT (Forster & Chambers, 1973; Frederiksen & Kroll, 1976; Hudson & Bergman, 1985; O'Regan & Jacobs, 1992; Weekes, 1997), accuracy in PITs (Aghababian & Nazir, 2000; Nazir, et al., 1998), or refixation probability and gaze duration in eye movement studies (McConkie et al., 1989; O'Regan et al., 1984). Lately, even in electrophysiological studies peak latency of early negativity has been found to be correlated with word length (Osterhout et al., 1997). Sometimes, even the most reliable effect in visual word recognition research, the word frequency effect, may be driven by the confound with word length (see Hudson & Bergman's, 1985, reanalysis of Forster and Chamber's data, discussed in the above chapter). However, the existence of an independent word length effect for words has also been questioned. Recently, Weekes (1997) found in a NT that the word length effect obtained for low-frequency words no longer existed when neighborhood size, number of friends, and average grapheme frequency were included in the regression analysis. Only for nonwords, Weekes found a significant contribution of the number of letters in his regression study. However, my hypothesis is that failures to find word length effects are not only due to unfavorable confounds with other variables. Instead, I suggest that null effects of word length can sometimes be due to the presentation of words to optimally skilled readers under optimal presentation conditions and their investigation using measures of central tendencies in the general linear model.

Evidence for my hypothesis comes from the results obtained with non-optimally skilled readers and the interactions of the word length effects with skill. In normal readers, skill interacts with word length in the way suggested above (Butler & Hains, 1979): Skilled readers – as measured by a vocabulary test – clearly showed less of a word length effect in both, the NT and the LDT. Considered that the participants are normal students in an introductory course the size of the interaction is surprising. While high-vocabulary participants responded about 100 ms faster for 2-4 letter words this advantage almost linearly increased with every additional letter, until the advantage for the high vocabulary group was about 300 ms for 12 –14 letter words. In the LDT, the interaction was disordinal over different word lengths for the

skill groups. While for short words surprisingly the unskilled group was faster, for longer words the skilled group responded faster. However, in both experiments, skill (large vocabulary) clearly reduced the word length effect. This effect for normal readers converges with results from beginning readers. Aghababian and Nazir (2000) showed that the word length effect becomes reduced as reading instruction proceeds. Perfetti et al. (1979) directly assessed reading skill of young readers and found systematic interactions of reading skill with word length. Word length affected unskilled readers more than skilled readers in their experiments and particularly so, when the word was presented without context.

Finally, evidence for the hypothesis that word length effects are particularly likely to occur in unskilled readers comes from pure alectic readers. There, the word length effect is considered to be the hallmark effect of pure alexia (Behrmann et al., 1998; Montant et al., 1998). Alectic patients usually show huge word length effects (see Behrmann et al., 1998, for a review). Based on computational evidence, Montant et al. (1998) suggest that in their patient a reduced ability to use top-down information may hinder the processing of visual input information particularly for poor viewing positions, and may consequently be also responsible for an increased word length effect in alectic patients. The interaction of word length and imageability in the patients of Behrmann et al. (1998) is also in line with the assumption of reduced top-down effects. In alectic readers word length effects are particularly big when imageability is low.

When stimulus attributes or task demands are difficult the word length effect is also more likely to occur. Hudson and Bergman (1985) found in the LDT that when word-like pseudowords were used a significant length effect could be obtained (Experiments 1 and 2). In contrast, they did not find a word length effect when word-unlike nonwords were used, no matter if they were considered orthographically legal or illegal (Experiments 3 and 4). With regard to visual attributes, Koriat and Norman (1985) investigated the interaction of word length effects with disorientation of a given word. They obtained stronger word length effects when the disorientation was stronger. Again, in orientations in which word recognition was most difficult the word length effect was strongest. Finally, Perfetti et al. (1979) found that word length effects tended to be stronger in young readers when no context was given for target word recognition.

In sum, there seems to be converging evidence that word length effects are stronger when stimulus conditions, presentation conditions or participants' skills are non-optimal. In particular, word length effects tend to increase with poor viewing position and low frequency. These interactions are of direct relevance for the hypotheses guiding my experiments. Therefore, they will be reviewed in more detail below.

4.3. Interactions with Viewing Position and Frequency

Recognition of longer words suffers more from non-optimal viewing positions than recognition of shorter words. Most prominently, this has been investigated with stimulus projections into different hemispaces or hemifields. Hemifield interactions with word length seem to occur even more consistently than with word frequency. Recognition is more hindered by word length, when a word is presented in the left hemifield and, initially projected into the right hemisphere than vice versa (e.g., Iacoboni & Zaidel, 1996; see Nazir, 2000, for a review). When viewing position is continuously manipulated, word length effects tend to be larger on non-optimal viewing positions in the LDT, the NT (O'Regan & Jacobs, 1992), and in PITs (Aghababian & Nazir, 2000; Nazir et al., 1998). In eye movement studies, however, the interaction between viewing position and word length is not that clear (O'Regan et al., 1984; Vitu & O'Regan, 1988). Like for frequency effects, one reason may be that in continuous text reading the matter is complicated by the controversy of how many words are processed and preprocessed and how that influences word recognition attributes (cf. Deubel et al., 2000).

Yet, one important caveat has to be made when these effects are discussed. While normally word length effects are simply compared with respect to central viewing position (i.e., a relative letter position, that is a function of word length) it is not clear, which non-optimal viewing positions should be compared. Therefore, in ANOVAs also relative viewing positions are compared, i.e., each word is divided into a number of equally sized zones and the respective viewing positions are in the middle of those zones (O'Regan & Jacobs, 1992). Thus, the relative viewing position was identical across stimulus groups. One could also analyze absolute viewing positions to examine word recognition effects when fixation is at the first, second, or last letter of a word, no matter what the length is. Sometimes (e.g., O'Regan et al., 1984), a mixed procedure is used: From the central viewing position, i.e., a relative letter position that is a function of word length, absolute letter positions are computed to the left or the right from central fixation. Then, a non-optimal viewing position for a long word may be not too far from the middle of the long word while it is on or even off the end or beginning of a short word. Clearly, when the non-optimal viewing position is relatively more towards the middle for long words, an interaction of word length with viewing position is hardest to detect (see O'Regan et al., 1984; O'Regan & Jacobs, 1992; Vitu & O'Regan, 1988, for graphics illustrating that matter). I propose that one should either stay with relative positions or use absolute letter positions to compare different word length. Therefore, a word length interaction with viewing position should not generally be rejected if this rejection is only based on that very combination between relative and absolute viewing position most likely producing null effects. Therefore, for the

analyses in this study, relative viewing position will be used. However, an extra graph will be given to allow for a direct comparison of absolute viewing position distance from (relative) central fixation position.

Not only main effects of word length (and frequency) themselves increase under non-optimal conditions but sometimes also their interaction. A typical example for such a study is again that of O'Regan and Jacobs (1992) who found word length by frequency interactions in both, the NT and the LDT. In the LDT, the interaction was in the expected direction: The frequency effect increased with word length, and vice versa the word length effect was larger for low-frequency words (about 18 ms per letter) than for high-frequency words (about 10 ms per letter). The above interaction was observed for global means over all viewing positions. However, it seemed to vary with viewing position. The interaction between word length and frequency for the middle and optimal positions seemed to be much weaker than for the other positions (see pp. 188-189). Similarly, in the NT, the interaction between frequency and word length was significant when all viewing positions were considered but it was reduced to a marginally significant effect when only the central viewing position was considered (p. 193). Based on these results, O'Regan and Jacobs (1992, p. 196) *"raise the question of which is the correct position when measuring the effects of length and frequency."* Convergent results are also obtained by Perfetti et al. (1979). Not only did word length and frequency interact in the expected direction in some of their NT experiments (Experiments 2 and 3) but they interacted particularly for the unskilled young readers. Unskilled readers particularly seemed to suffer when words were long and low-frequent compared to skilled readers and when no context was provided. Finally, in a word reading task, where participants just had to press a button when they had read the word, the same word length by frequency interaction has been found (Lorch, 1986). In contrast, in eye movement studies additive effects were reported, when both, length and frequency were investigated (McConkie et al., 1989; Vitu et al.; 1990). In sum, this review indicates that the word length effect itself and its interaction with frequency increase under inferior (visual) conditions.

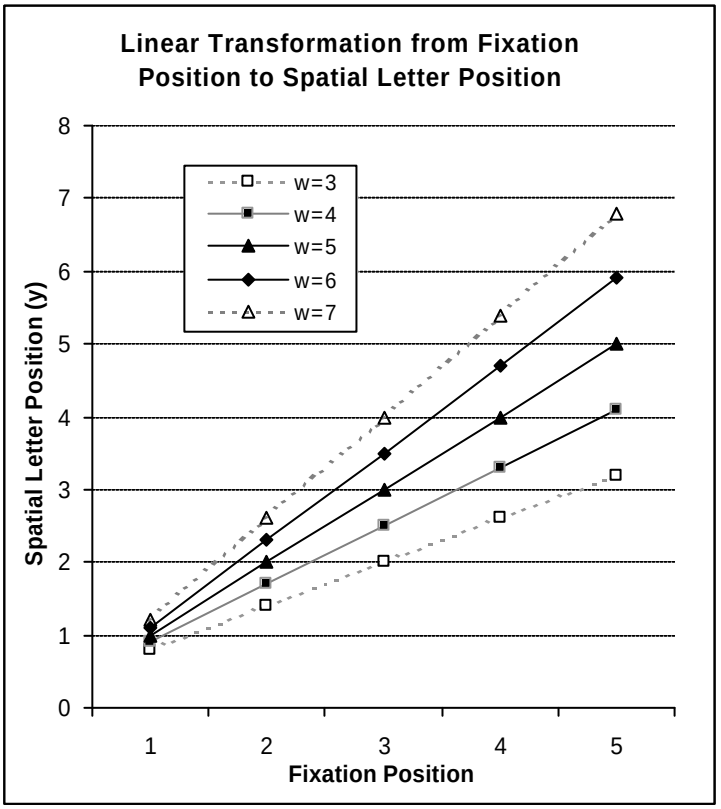
4.4. Generalizing the Model to any Word Length and Number of Fixation Positions

What does this underadditivity imply for the specific form of the viewing position curves? The model of Nazir et al. (1998) does not yet allow an answer to this question because it does not allow fits for all possible word lengths when the number of fixation positions is held constant between word lengths. Therefore, I generalized the model by simply assuming that the linear drop-off in legibility is not discrete but continuous when fixation is not exactly on a certain letter position but somewhere in between. Following the principle of nested modeling (Jacobs & Grainger, 1994), the old model is a special case of

the generalized model, which will be demonstrated below. The generalized model will enable me to fit empirical data for any word length and any number of fixations. Consequently, its architecture will allow for testing more precise hypotheses about the interaction of word length and viewing position and the alteration of the word length by frequency interaction by viewing position.

Let w be the word length and n_f the number of fixation points used. Let us further assume that the *i'th* $(i = 1,...,w)$ letter position is at the center of the *i'th* letter and let f $(f = 1,..., n_f)$ be the location of fixation, with the leftmost location being $f = 1$ and the rightmost $f = n_f$. For words with $w \neq n_f$, each word shall be divided into n_f equally-spaced zones and fixation is in the middle of these zones. Then, fixation is not exactly on the center of each letter anymore but somewhere in between. By a simple linear interpolation I can compute the spatial letter position y for each fixation position f for the case $w \neq n_f$ by

$$y = \frac{w}{n_f} f + \frac{1}{2}(1 - \frac{w}{n_f}).$$



Equation 2 and Figure 2: Linear transformation from fixation position to spatial letter position. If the word length w does not equal the number of fixation positions n_f, then the fixation position f is not exactly anymore at the center of the letter i, but somewhere to the right or the left of this center. The transformation is such that the center of letter i is on spatial letter position y = i. This is achieved by assuming that the first letter extends from spatial letter position 0.5 to 1.5 with the center of this area being 1. Note that this construction implies that when fixation is on the center of the first letter, for $w = n_f$ it holds that $y = f = 1$ and not $y = 0.5 \neq f$ (see Figure 2, in which $n_f = 5$, for a word length $w = 5$)[5]. Figure 2 illustrates the linear transformation of Equation 2 for different word lengths (see text for details).

[5] One could also assume that $y = 0.5$ and that the first letter extends from spatial letter position 0 to 1. However, this convention has no direct consequences for the model fit. I chose the first possibility to follow the principle of nested modeling because fixation on the first letter in the model of Nazir et al. (1998) means also $y = f = 1$. Besides, the following equations can also be simpler expressed with the chosen convention.

To better understand Equation 2, consider the following example: For a three-letter word (*w = 3)* and five fixation positions (n_f *= 5),* each fixation zone extends to 3 / 5 = 0.6 letter widths. The first (leftmost) fixation position *f = 1* is then in the center of the leftmost fixation zone, i.e. 0.3 letter widths from its left edge. The left edge of the first fixation zone starts at the left edge of the first letter, i.e., at spatial position 0.5 (see Equation 2). Thus, the first fixation is at spatial letter position y = 0.5 + 0.3 = 0.8, i.e., a little left from the center of the first letter (which is at y = 1). The other spatial letter positions of the fixation points are then in a distance of 0.6 letter width (the width of one fixation zone), i.e., at 1.4, 2.0, 2.6, 3.2. The illustration of this example is the lowest curve in Figure 2.

$$p_{A,d}(i, r_f, y) = \begin{cases} r_f - (i-y)d & : \quad y < i \\ r_f & : \quad y = i \\ r_f - (y-i)Ad & : \quad y > i \end{cases}$$

<u>Equation 3:</u> *For this equation, let d be the drop-off rate to the right and let A be the asymmetry ratio of the drop-off rates to the left and the right. Then Ad is consequently the drop-off rate to the left. According to the model the recognition probability* $p_{A,d}(i, r_f, y)$ *of a letter i is assumed to decrease linearly with its spatial letter distance from fixation y. It is important to understand that the integer i is the ordinal position of a given letter while the rationale number y is a spatial position in letter space. Equation 3 works as easy as given above because Equation 2 constructed the spatial letter positions such that the center of letter i is at the spatial letter position y = i. In the case y < i, the letter i is to the right from fixation f and its probability of recognition* $p_{A,d}(i, r_f, y)$ *decreases linearly with the distance i - y and the drop-off rate d from the recognition probability at fixation* r_f*. In the case y = i the letter i is direct at the fixation f and its probability of recognition is* $p_{A,d}(i, r_f, y) = r_f$*. Finally, in the case y > i the letter is to the left from fixation f and its probability of recognition* $p_{A,d}(i, r_f, y)$ *decreases linearly with the distance y - i and the drop-off rate Ad from the recognition probability at fixation* r_f*.*

More conveniently this can be expressed as:

$$p_{A,d}(i, r_f, y) = \begin{cases} r_f - (i-y)d & : \quad y \le i \\ \\ r_f - (y-i)Ad & : \quad y > i \end{cases}$$

<u>Equation 4:</u> *Equation 4 is a just shorter, but less transparent version of Equation 3 which was constructed parallel to Equation 1 to enable convenient comparisons.*

This enables now the step from computing the single letter recognition probabilities to computing the word recognition probability.

$$PW_{A,d}(r_f, w, y) = \prod_{i=1}^{w} p_{A,d}(i, r_f, y)$$

<u>Equation 5:</u> The probability PW for recognizing a word is just the product of the independent recognition probabilities $p_{A,d}(i, r_f, y)$ for its constituent letters i, with i ranging over the whole word length w from i = 1, … ,w. A and d denote the free parameters of the parametric family of probability distributions $PW_{A,d}$, while r_f, w, and y denote the fixed parameters which are function variables of $PW_{A,d}$.

By inserting $p_{A,d}(i, r_f, y)$ into Equation 5 I obtain (with special cases for fixation on the first and on the last letter).

$$PW_{A,d}(r_f, w, y) = \begin{cases} \prod_{\{i \in \aleph : y \le i \le w\}} (r_f - (i-y)d) & : \quad y = 1 \\ \prod_{\{i \in \aleph : y \le i \le w\}} (r_f - (i-y)d) \prod_{\{j \in \aleph : j < y\}} (r_f - (y-j)Ad) & : \quad 1 < y < w \\ \prod_{\{j \in \aleph : j < y\}} (r_f - (y-j)Ad) & : \quad y = w \end{cases}$$

<u>Equation 6:</u> Equation 6 is just the insertion of Equation 3 into Equation 5. For better convenience, the integer i = 1, ..., w denotes the letters right from spatial fixation position y (y <= i) and the integer j = 1,..., w, denotes the letters right from spatial letter position y (j < y).The conditions at the product terms P describe just the above sentence mathematically.

However, the spatial letter position y is a function of the fixation *f*, the word length *w* and the number of fixation points n_f (Equation 2). If I resolve this functional relation, I get the final raw equation:

$$PW_{A,d}(f, n_f, r_f, w) =$$

$$\begin{cases} \prod_{\left\{i \in \aleph : \frac{w}{nf} f + \frac{w}{2}(1 - \frac{w}{nf}) \le i \le w\right\}} (r_f - (i - \frac{w}{n_f} f + \frac{1}{2}(1 - \frac{w}{n_f}))d) & : \quad f = 1 \\ \prod_{\left\{i \in \aleph : \frac{w}{nf} f + \frac{1}{2}(1 - \frac{w}{nf}) \le i \le w\right\}} (r_f - (i - \frac{w}{n_f} f + \frac{1}{2}(1 - \frac{w}{n_f}))d) \bullet & \\ \prod_{\left\{j \in \aleph : j < \frac{w}{nf} f + \frac{1}{2}(1 - \frac{w}{nf})\right\}} (r_f - (\frac{w}{n_f} f + \frac{1}{2}(1 - \frac{w}{n_f}) - j)Ad) & : \quad 1 < f < w \\ \prod_{\left\{j \in \aleph : j < \frac{w}{nf} f + \frac{1}{2}(1 - \frac{wl}{nf})\right\}} (r_f - (\frac{w}{n_f} f + \frac{1}{2}(1 - \frac{w}{n_f}) - j)Ad) & : \quad f = w \end{cases}$$

<u>Equation 7:</u> Equation 7 is just the insertion of Equation 2 into Equation 6: y is substituted by the right term of Equation 2. It is the final raw equation of the generalized model.

One can easily see that Equation 7 is just a generalization of Equation 1. When the number of fixation points equals word length (i.e, $w = n_f$) it holds that:

$$\prod_{\left\{i\in\aleph:\frac{w}{nf}f+\frac{1}{2}(1-\frac{w}{nf})\le i\le w\right\}}(r_f-(i-\frac{w}{n_f}f+\frac{1}{2}(1-\frac{w}{n_f}))d)$$
$$\overset{w=nf}{=}\prod_{\{i\in\aleph:f\le i\le w\}}(r_f-(i-f)d)$$
$$\overset{k=i-f}{=}\prod_{\{k\in\aleph:1\le k\le w-f\}}(r_f-kd)$$

<u>Equation 8:</u> This equation proves the principle of nested modeling for the generalized model. For w = n_f is w / n_f = 1 which makes the model identical to Equation 1 (except for the distance k = i – f which is directly incorporated in Equation 1).The prove is accordingly for the other product term.

4.5. Objectives and Specific Hypothesis Derived from the Generalized Model

Because of its architecture the model already predicts a strong word length effect on all viewing positions for simple visual reasons. Consider, for example, the case for central fixation position in five- or seven-letter words. The recognition likelihood of the inner five letters of the seven-letter word does - with drop-off rate and asymmetry ratio being equal for both word lengths - match the recognition likelihood for the five letter word. However, if the drop-off rate is equal for both word lengths, the seven-letter word should be recognized less accurately because two more letters (the first and the seventh letter) are considered by the model. Their letter legibility probabilities are multiplied with the product of the probabilities of the inner five letters. If drop- off rate is greater than zero, the letter legibility probabilities are smaller than 1 and, thus, the overall product for the seven-letter word is smaller than the product for the inner five letters. Moreover, the word length effect should also be particularly pronounced for poor viewing positions. Consider the middle and the final viewing position for each word. The model predicts that the likelihood of recognizing a letter decreases with distance from viewing position. Because all likelihoods are multiplied in the current version of the model any single very low recognition probability is particularly detrimental.

Consider the following example:

Let $0 < a,b < 1$ and $c := \frac{a+b}{2}$

Then: $c^2 = \left(\frac{a+b}{2}\right)^2 = \frac{a^2+2ab+b^2}{4} \overset{a^2+b^2\ge 2ab}{\ge} \frac{4ab}{4} = ab$

Thus, the likelihood of recognizing a word with two equal letter legibility probabilities *c* is greater than recognizing a word with one high and one low probability (*a* and *b*). Essentially, this illustrates that a very

low legibility probability of any letter is particularly devastating for recognition of the whole word, even if the legibility of the other letters is relatively high.[6]

The longer the word is, the greater becomes the distance of the least legible letter from fixation, the smaller becomes the recognition probability of this letter, and, finally, the more overall recognition performance deteriorates in the model. This principle is one reason for the optimal viewing position being mostly in or slightly left from the middle. However, it also predicts that initial or final fixation for long words lead to particularly bad results because the most distant letter(s) are then even more distant than for short words. Thus, their influence should be even more detrimental. In sum, the word length effect itself and the *qualitative* nature of the hypothesized interaction of word length and viewing position would be captured by the model as it is. However, it is not clear, whether the model is *quantitatively* able to catch this variation of word length and its interaction with viewing position with relatively invariant parameters for different word length conditions.

The model has yet no parameter to represent the influence of perceptual frequency of location for different word lengths and the interaction of word length with lexical frequency. Nevertheless, some predictions can be derived. First, the perceptual frequency hypothesis of location of Nazir et al. (1998) predicts that a word is recognized best at that location where it was previously observed most frequently. If this hypothesis proves to be true, the asymmetry ratio should change as a function of word length. The viewing position curve for long words, whose most frequent viewing position is more shifted toward the left should be fitted with a higher asymmetry parameter than short words. In contrast, if the perceptual frequency hypothesis is wrong no such alteration should be observed. However, if such an alteration is observed, the visual asymmetry parameter would code two things: i) the visual asymmetry of recognizing single letter in strings of any length and ii) the shift in the optimal viewing position for a whole word depending on where this word had been seen in the past. Future model refinements should disentangle these two processes if possible to allow independent manipulations of individual letter legibilities and word length. Second, the interaction between word length and frequency should lead to a greater frequency improvement for long words on poor viewing positions. Because the drop-off parameter also determines the form of the curve and not only the height, the curves form should particularly differ for high- and low-frequent long words: It should be much flatter for long, high-frequent words. Although, this alteration is exactly what would be qualitatively expected, it is not clear whether the model captures these

[6] The above equation proves the simple fact, that a square is the rectangle with the greatest area inside, when circumference is held constant.

curve forms and their alterations quantitatively well for long words. It may be that the direction of alteration is predicted well by the model, while the specific fits for the empirical data are poor.

I will take advantage from the attribute of mathematical and algorithmic models (Jacobs & Grainger, 1994) that they - in contrast to verbal models - provide quantitative model fits or simulations for empirical data. In Study 2, the generalized model will be tested not only qualitatively but also quantitatively.

5. On the Visual Alteration of Orthographic-Lexical Neighborhood Effects

The first chapters were mainly concerned with individual item effects, the perceptual frequency effect, the lexical frequency effect and the word length effect. The investigation of orthographic neighborhood effects and their alterations by visual attributes represents a further generalization step. Neighborhood effects do not so much refer to individual properties of a given word but to its relations with the rest of the lexicon. Orthographic neighborhood by its original definition is a similarity index in orthographic-lexical space, restricted, however, to the words of the same length as the target. However, I will present evidence that understanding orthographic neighborhood as a solely orthographic-lexical variable is not sufficient. The visual dimension is lacking, when orthographic neighborhood is regarded as a similarity index in orthographic-lexical space. Cutting out any given orthographic-lexical hyperspace (with fixed visual conditions) from visual-orthographic-lexical space is not sufficient as any such hyperspace is not representative for other parallel hyperspaces with differently fixed visual conditions.

5.1. Introduction and Definition

Orthographic neighborhood density (Coltheart, Davelaar, Jonasson, & Besner, 1977) is probably the second most popular index in the visual word recognition literature after word frequency (Forster & Chambers, 1973), but certainly an index whose effect is much more controversial (e.g., Andrews, 1997; Grainger & Jacobs, 1996; Sears, Hino, & Lupker, 1995; Snodgrass & Mintzer, 1993; Ziegler & Perry, 1998). The definition of orthographic neighborhood density (N) is usually adopted from Coltheart et al. (1977): An orthographic neighbor is any word that can be created by changing one letter of the stimulus word while preserving all other letter positions.

N is controversial and interesting because it has been found to be either facilitatory or inhibitory (for reviews see Andrews, 1997; Grainger & Jacobs, 1996), whereas other relevant variables have consistently produced facilitatory (e.g., frequency) or inhibitory effects or null effects (consistency,

regularity, e.g., Glushko, 1979; Jacobs et al., 1998; Jared et al., 1990; Plaut, et al., 1996; Ziegler et al., 1997). Numerous determinants of facilitatory or inhibitory N effects have been proposed, the most important being the task (e.g., Andrews, 1997; Grainger & Jacobs, 1996), the language (e.g., Andrews, 1997; but see Perea & Pollatsek, 1998; Van Heuven, Dijkstra, & Grainger, 1998; Ziegler & Perry, 1998), the relationship between the frequency of the neighbors and the target word (e.g., Carreiras, Perea, & Grainger, 1997; Forster & Shen, 1996; Grainger, 1990; Grainger & Jacobs, 1996; Grainger, O'Regan, Jacobs, & Segui, 1989; Huntsman & Lima, 1996; Sears et al., 1995), the dependent variable (e.g., Andrews, 1997; Snodgrass & Mintzer, 1993; Ziegler et al., 1998), or the phonological properties of the neighbors (e.g., Colombo, 1986; Forster & Davis, 1991; Forster & Taft, 1994; Graf & Nuerk, 1999; Peereman & Content, 1997; Ziegler & Perry, 1998). The picture is further complicated by interactions between the numerous determinants as, for example, the number of higher frequency neighbors and task specificity (e.g., Andrews, 1997; Grainger & Jacobs, 1996). However, despite conflicting results and various attempts to resolve these conflicts (e.g., Andrews, 1997; Grainger & Jacobs, 1996; Ziegler & Perry, 1998), one common (implicit or explicit) assumption is made in virtually all of these papers: That it suffices to understand neighborhood effects as orthographic-lexical (or phonological-lexical) effects[7]. For example, Andrews (1997) discusses neighborhood conflicts as the effect of orthographic similarity on lexical retrieval. In a similar way, Ziegler et al. (1998) interpret the neighborhood density effect in the screen fragmentation task within an interactive activation framework as a top-down effect compared to a bottom-up effect of letter confusability. In this thesis, I will argue that it is not sufficient to regard neighborhood effects as (lexical) top-down effects. Instead, at least in PITs N effects may better be understood as interactive visual-orthographic-lexical effects, whose simple interpretation as indices for similarity in orthographic-lexical space falls short of accounting for the data.

5.2. The Interaction of Orthographic Information with Visual Determinants

My assumption that the presumed effects of orthographic-lexical neighborhood are better understood as visual-orthographic-lexical effects is stimulated by evidence that orthographic information indeed interacts with visual determinants of viewing position. One line of evidence comes from eye movement studies that have shown that viewing position interacts with the uniqueness information provided by the

[7] Sometimes the confound of N with sublexical orthographic or phonological indices (rather than lexical indices) is emphasized (e.g., in Frauenfelder, Baayen & Hellwig, 1993, the confound between the orthographic index bigram frequency and N; or in Ziegler & Perry, 1998, the confound between a sublexical phonological index and N). For the current argument this distinction is not important because visual aspects of sublexical units are not emphasized in these accounts either. I will, however, come back to this distinction in the general discussion and the discussion of incorporation of the new model.

orthographic structure of the word. Early evidence comes from O'Regan et al. (1984) who investigated gaze duration in first and second reading of 9-12 letter words that are *informative* either at the beginning or at the end of the word. *Informative* in their study meant that there existed no word in the French lexicon that had either the same six letters at the beginning or the same six letters at the end. They found an interaction of the location of this lexical uniqueness information in the word and viewing location in the expected direction. Gaze duration was shorter when viewing position was on the informative part of the word (O'Regan et al., 1984). This pattern of results was not restricted to gaze duration, as Brysbaert et al. (1996) found the same pattern of response in a PIT. Information in words also seems to play a role in parafoveal preview processing of words during text reading. Beauvillain et al. (1996) showed that the initial landing position in words with informative, yet orthographically irregular beginnings were more to the left than in words with uninformative beginnings (see also Underwood et al., 1990, for similar results; but see Rayner & Morris, 1992, for null results). Taken together, these results may suggest that the distribution of orthographic information attenuates VPEs in foveal and parafoveal vision in a characteristic way.

However, VPEs are not the only effects suggesting that orthographic information interacts with visual determinants. Normally, orthographic neighborhood is defined in terms of a position-specific exchange of a letter. Words differing in two letter positions should be orthographically less similar. However, transposition neighbors (in which the order of two letters is switched rather than one letter exchanged) that differ in two positions produce inhibitory effects in both, the LDT and the NT (Andrews, 1996; Chambers, 1979): This finding, particularly in the NT, contrasts with findings that N effects in the NT are normally facilitatory (Andrews, 1997). A possible account is that visual confusability of transposition neighbors is very high due to spatial or positional uncertainty of the letter order in word recognition and that this high confusability leads to inhibitory effects on performance. Finally, evidence that letter confusability exerts its effects on orthographic-lexical retrieval comes from all experiments validating the AVM of Paap et al. (1982) because the basis of the AVM is a position-unspecific letter confusability matrix that determines which words are selected for the response set (Paap & Johansen, 1994).

Thus, there is evidence that visual determinants, viewing position, or letter confusability exert their effects on the orthographic-lexical information distribution within words but to my knowledge there is only one study, that has studied the effect of visual determinants on neighborhood effects directly. Grainger, O'Regan, Jacobs and Segui (1992) manipulated fixation position (either letter position 2 or 4), neighborhood position (either letter position 2 and 4), and the number of higher frequency neighbors in

two LDTs. In the literature, it is mostly assumed that there is no main effect of neighborhood position and that this variable, hence, can be neglected in neighborhood models (Forster & Taft, 1994). While Grainger et al. (1992) found no general neighborhood position effect, they found, among other interactions, an interaction of neighborhood position with viewing position. For words with higher frequent neighbors on the second position, fixation position on the critical letter facilitated lexical decision latency. In contrast, for words with neighbors on the fourth position, no such effect of fixation position was found. When an interference measure (neighborhood frequency effect[8]) was computed, it seemed to be lower when fixation was on the disambiguating letter (see Grainger et al., p. 54) rather than off the disambiguating letter. Grainger et al. (1992) interpreted their finding as an indication for an interaction between visual and lexical factors: "*These results strongly suggest that on the one hand, the effects of competition between lexical representations can be modified by the relative visibility of the stimulus word's letters and, on the other hand, that the distribution of lexically constraining information across the stimulus word influences the consequences of varying the relative visibility of the component letters*."

While I generally agree with the conclusions of Grainger et al. (1992), their experiment has certain limitations. First, their conclusions that visual factors influence the effect of positional higher frequent neighbors basically rests on an interaction present only for position 2 but not position 4 in five-letter words. Possible implications for other viewing positions, particularly for the central viewing position mostly used in word recognition experiments can only be assumed and have to my knowledge not been studied yet. Will central fixation position 3 (in five-letter words) interact with neighborhood position like fixation position 2 or will there be no effect for position 4? Moreover, we do not know how systematic and general this asymmetrical effect of viewing position by neighborhood distribution is yet as only two letter positions have been examined. Second, while the position of high-frequency neighbors was controlled (positions 2 or 4), position of the other neighbors appears not to have been controlled. In this respect, the results for the control words without any higher frequency neighbor at any position are particularly remarkable. The control words for words with a higher frequency neighbor on the fourth position (short HFN4 words) were better recognized *off fixation* on letter position 2 than *on fixation* on letter position 4 (37 ms in Experiment 1 and 24 ms in Experiment 2). In Experiment 2, a facilitatory effect for control words for HFN2 words was found for fixation on position 2. However, surprisingly for HFN2 words themselves, no effect of fixation position was found in Experiment 1. Thus, there are some inconsistencies in the data. The control words

[8] The neighborhood frequency effect refers to the fact that higher frequent neighbors (and in particular exactly one higher frequent neighbor) tend to produce stronger inhibitory effects than other neighbors (cf. Grainger et al., 1989; Grainger & Jacobs, 1996).

seemed to have some attributes that sometimes interacted with fixation position even though the words with higher frequent neighbors did not show that same interaction all the time. A possible account is that the position of lower frequency neighbors, known to have an inhibitory influence in the PIT (unless sophisticated guessing is allowed, cf. Andrews, 1997) and an inconsistent influence in the LDT, also played a role in their experiment. Third, the visual manipulation was restricted to having higher frequency neighbors on and off fixation. It was not examined whether the distance to fixation played a role but this factor must be considered because the legibility of the critical letters systematically drops with increasing distance to fixation (Nazir et al, 1992). Fourth, the findings were restricted to the LDT. Due to the principle of functional overlap it is worthwhile to know whether a certain effect is task-specific or task-independent.

Study 3 was constructed to tackle these issues. I manipulated viewing position more continuously than on and off fixation and did not restrict the stimulus selection to words with neighbors on certain positions but not on others. No such strict criterion was incorporated. Instead, I selected words that had most of their neighbors (and thus their information ambiguity) at the beginning and compared them with words that had most of their neighbors at the end. The specific measure I developed to construct the respective stimuli groups, is given in the introduction to Study 3. For the hypotheses concerning the model of Nazir et al. (1998) it suffices to know that information ambiguity as indexed by neighborhood density distribution was either towards the beginning or towards the end of a word.

5.3. Hypotheses Concerning the Model of Nazir et al. (1998)

In the model of Nazir et al. (1998) the asymmetry parameter is indexed as a purely visual parameter. The asymmetry ratio in the drop-off rate stems from the asymmetry in letter legibility found in empirical experiments (Nazir et al., 1992). However, the following considerations might indicate how the visual asymmetry ratio might change as a function of orthographic-lexical manipulations.

If N is better understood as a visual-orthographic-lexical measure, then it should interact with letter legibility in a characteristic way. If the information ambiguity (i.e., many neighbors) is rather at the beginning of a word, then it should be relatively better for word recognition performance to fixate at or near the beginning because then information ambiguity is exactly at those letter positions which are best legible according to the model. In contrast, if information ambiguity is at the beginning and fixation is near the end of the word then these fixation positions should be even more detrimental than normal because information ambiguity is at badly legible letter positions. If the information ambiguity (i.e., many neighbors) is at the end of the word, the pattern should be reversed. Then fixation on the beginning should yield

relatively worse performance and fixation near the end relatively better performance. This implies a shift in the asymmetry of the curve and the asymmetry ratio parameter in the model of Nazir et al. (1998). If the information is at the beginning of the word, the curve should consequently even be more asymmetrical than normal while it should be more symmetrical when most neighbors are at the end of the word because in this case relatively better performance would be expected at the end of the word and relatively worse at the beginning. If the information is at the end of the word, the viewing postion curve should – vice versa, become more symmetrical.

If these considerations are correct, the hypotheses concerning the model are straightforward. If the asymmetry ratio is a purely visual parameter in the current model, it should not change with neighborhood distribution in the way discussed above. Neighborhood density would then represent an orthographic-lexical measure, that at least in this task does not influence the visual asymmetry parameter. However, if the above hypotheses hold, the curves should shift in asymmetry as a function of neighborhood distribution. The notion that the asymmetry parameter in the current model is a purely visual parameter would be wrong because it could be influenced by lexical information ambiguity. Then a refined model would have to take into account local information ambiguity as indexed by neighborhood distributions.

If an interaction between neighborhood distribution and viewing position would be observed, then neighborhood size would be better understood as a visual-orthographic-lexical measure. In this case, the visual-orthographic-lexical account of neighborhood will be validated with another visual manipulation in another task to investigate whether it is task-specific or manipulation-specific. The specific manipulation and its incorporation, however, will be introduced before the experiment designed for that validation.

6. Summary of the Introduction

This review aimed at pointing out the importance of visual determinants of visual word recognition. It should raise doubts that the interpretation of null effects as strong evidence against the role of visual determinants is wholly conclusive. I tried to raise such doubts in three domains. First, I suggested that item-specific perceptual frequency of visual word form may still play a role in visual word recognition. This claim is in contrast to the literature. I tried to show that the conclusion that visual word form did not play any role in visual word recognition rests on a logic that may not always be appropriate. Specifically, null interactions in an ANOVA based on the general linear model do not need to be deterministically interpreted using AFL. Instead, additive effects in measures of central tendencies can also be described by interactive activation models assuming interactions between lexical and perceptual processes.

While for perceptual frequency I suggested that - contrary to the literature - a specific visual effect does exist, I hypothesized in two other domains that even the most established hallmark effects of visual word recognition are altered by visual determinants in a specific non-linear way. Thus, examining these effects using the principle of isolated variation may produce results which cannot be generalized to other visual conditions. More specifically, I asked whether null interactions of measures of central tendencies in an ANOVA between visual determinants and lexical frequency can be interpreted using AFL following the general linear model. I argued that the frequency effect and its interactions with other variables increase as presentation conditions, stimulus attributes and participants' skills become poorer. Examining the frequency effect and its interactions under near-optimal conditions may lead to their underestimation. In particular, if null interactions of frequency with other variables, such as word length, are obtained under near-optimal conditions, these results cannot be generalized to other visual conditions. Because I will investigate the interaction of frequency with word length, I also reviewed the literature concerning the word length effect and its interactions. Similarly to the frequency effect, I argued that the size of the word length effect may be underestimated and its existence may not be detected under near-optimal conditions (e.g., central unlimited fixation in skilled students). Therefore, in particular, for word length by frequency interactions, null effects under near-optimal visual conditions may be observed, that do not generalize to poorer visual conditions.

In the third domain, I examined the visual determinants of orthographic neighborhood effects. While perceptual and lexical frequency are individual attributes of a given item, orthographic neighborhood directly indexes the relations of a target word to other words in the lexicon. Neighborhood effects are mostly interpreted as indexing relations in orthographic-lexical space, either orthographic-lexical competition or facilitation. I proposed that neighborhood is better understood as an index of similarity in visual-orthographic-lexical space. Experiments using the principle of isolated variation with constant, yet near-optimal presentation conditions (e.g., central fixations) may produce results that cannot be generalized to other presentation conditions.

For all these effects, hypotheses about the parameter changes in the model of Nazir et al. (1998) were derived. If perceptual frequency plays a role, the drop-off rate between different word forms should be the lower, the higher the perceptual frequency of that word form is. If perceptual frequency does not play a role, the drop-off rates should not differ much.

A little more complicated are the hypotheses concerning the frequency effect. If the frequency effect should be within the scope of a future mathematical viewing position model, a lexical parameter must be

added. However, it is not clear whether this parameter must be purely additive to the visual parameters or not. If frequency helps most under non-optimal conditions, the frequency effect should become larger for inferior viewing positions. Thus, not only the height of the viewing position curve should change but also its form. If visual manipulation and frequency do exert their influence on successive independent stages, the lexical parameter should be purely additive. Currently, drop-off rate determines height and curve form of the viewing position curve. However, only the specific fits can show whether the alteration of height and form is successfully captured by the model or whether the alteration of curve form is over- or underestimated. In this introduction, the model was generalized to fit data for any word length and number of viewing position. As concerns the word length by frequency interaction, the hypothesis is a generalization from the previous hypotheses about the frequency effect. Frequency should help most for longer words (i.e., the inferior word length condition) and poorer viewing position. While this alteration of curve form may qualitatively be caught by the model the quantitative aspect of the fits remain uncertain. The word length effect also offers another test of the perceptual frequency hypotheses of location, that complements the investigation of the perceptual frequency hypothesis of item-specific word form above. The distribution of landing positions in long words is more asymmetrical than in short words. Therefore, according to the perceptual frequency hypothesis of location, the asymmetry in the model of Nazir et al. (1998) should increase with word length, i.e., long words should produce more asymmetrical curves.

Finally, it should be tested whether the asymmetry parameter in the model of Nazir et al. (1998) is a purely visual parameter. If this is the case, the asymmetry parameter should not be altered in a systematic way by the distribution of orthographic information ambiguity within a word. Then, also, neighborhood could be further regarded as a similarity index of lexical-orthographic space. In contrast, if the asymmetry parameter differs between stimulus groups with different distributions of orthographic information ambiguity, then the asymmetry parameter would not be a purely visual parameter. In the latter case, neighborhood should better be regarded as a similarity index in visual-lexical-orthographic space.

The results of all three studies will provide the constraints necessary for the creation of a mathematical viewing position model at the end of this thesis. The goal of this model is not only to account for a visual effect such as the viewing position effect but also for the frequency effect, the word length effect, the neighborhood effects, and possibly the interactions of these variables with each other.

II. Study 1: On the Perceptual Frequency Hypothesis of Item-Specific Word Form

1. Introduction

The purpose of the first study is to show that the perceptual frequency of visual word form may play a role in visual word recognition. It has been argued that *"reading uses visually based letter clusters of the same size and case"* (Mayall, et al. 1997). As discussed in detail in the introduction, this and similar arguments against the influence of item-specific visual word form essentially was based on interpreting null interactions in an ANOVA in the framework of the general linear model using AFL. In contrast, I argued that the above conclusion Mayall et al. (1997) reached after a thorough investigation of case-mixing effects in different conditions, is a consequence of a item-specific perceptual frequency effect. For English words this does indeed imply that words are best recognized in same size, same case (in particular small case), but for some German words this is different. This translingual difference is helpful to disentangle the two hypotheses. In English, the perceptual frequency hypothesis of item-specific visual word form on one hand and the notion that words are recognized in same size and same case on the other hand, lead to the same performance predictions. In German, this is not the case.

2. Experiment 1: Testing the Perceptual Frequency Hypothesis with German Nouns in a PIT

2.1. Introduction

German provides a very simple direct test of the perceptual frequency hypothesis of item-specific visual word form against all general hypotheses of word form effects: Nouns are usually written in capitalized form, i.e., with the initial letter in upper-case (IUC condition). In titles and advertisements, they are sometimes completely written in upper-case (AUC: All upper-case condition), but almost never all in lower-case (ALC: All lower-case condition). According to the perceptual frequency hypothesis of item-specific word form German nouns should be best recognized in the IUC condition because they are perceived in this form most frequently. Performance in the AUC and ALC conditions, by contrast, should not differ because they are both quite infrequent. Alternatively, as reviewed in the general introduction, accounts neglecting item-specific word form would lead to different predictions. If words are best

perceived in same size, same case presentation, and more specifically if transletter and multiletter features of same size and case are the critical perceptual units (Mayall et al., 1997), the AUC and the ALC condition should yield superior performance because we are looking at different cases within one word in the IUC condition. If abstract single letter processing in normal-mixed case presentation is hindered by case-swapping, again the AUC and ALC conditions should produce more accurate responses. In contrast, if the greater physical distinctiveness of lower-case letters is decisive (Paap et al., 1984), then the conditions containing most lower-case letters (ALC and IUC) should provide better performance.

It has been claimed that the initial letters play a special role in visual word recognition (e.g., Forster, 1990; Inhoff & Tousman, 1990; but see Grainger & Jacobs, 1993, for an alternative view). So, if recognition accuracy in the IUC condition is superior, it may be so because the first letter is particularly well legible. In the IUC condition the initial letter is the only one in upper-case, thus larger than in the ALC condition, and less laterally inhibited than in the AUC condition. However, it is well known that the legibility of a letter decreases with increasing distance to fixation location (e.g., Nazir et al., 1992). Clearly, according to both the empirical observations and the model of Nazir et al. (1998) any improvement of legibility is particularly likely if letter legibility is poor, i.e., if the distance from fixation is particularly wide. Thus, if the difference in performance between the ALC and IUC conditions was only or mainly due to capitalization at the initial letter position, one would expect a particular improvement for fixations on the last letter positions. To investigate this alternative possibility I varied viewing position within a word according to a Latin Square design (cf. O'Regan & Jacobs, 1992) and fitted the model of Nazir et al. (1998) to examine whether the model can successfully fit the data when alternative assumptions are incorporated (Nazir et al., 1998; see below).

2.2. Method

2.2.1. Participants

15 psychology students of the Philipps-University Marburg participated in the experiment. All were native German speakers and had normal or corrected to normal vision. They received course credit for participation in the experiment.

2.2.2. Stimuli and Design

50 German lemmas from the CELEX-Database (Baayen, Piepenbrock, & van Rijn, 1993) with a mean log frequency of 2.02 per million were used in Experiment 1. All stimuli were five-letter German nouns (see Appendix A). Within each participant each word was presented in three different forms, the IUC, ALC and AUC form. Each word was shown at five viewing positions. The word was divided into five equally sized zones and the five viewing positions were in the middle of those zones. The 50 stimuli in each word form were subdivided into five lists, each consisting of 10 words corresponding to the five initial viewing positions. The attribution of each one of those 10-word sublists to one viewing position was counterbalanced across participants, following a Latin Square Design (O'Regan & Jacobs, 1992). In this way, each word in each word form was shown at each fixation position equally often across participants. The viewing positions of the different word forms of one word within one participant were always different to avoid a confound of fixation position and word identity that could serve as hint for word recognition.

2.2.3. Procedure

The experiment was run on an Apple Macintosh Powerbook 190CS. Participants sat approximately 30 cm from the screen. A trial began with a fixation point in the middle of the screen. Approximately 500 ms after the onset of the fixation point a stimulus was presented in Courier 20 font for 50 ms, immediately followed by a mask of seven hashmarks. From a distance of 30 cm the letters subtended a matrix of 0.95° height and 0.76° width. The participants' task was to type in correctly what they had seen. When the participant hit the return-key the next trial was started. Participants received 10 training trials in the training session and five buffer trials in the experimental session prior to the pseudorandomized 150 experimental trials. They were instructed to respond as accurately as possible. The whole experiment took about 20 minutes.

2.3. Results

Figure 3 shows clear main effects of viewing position ($F(4, 56) = 17.41$, $p < .001$) and stimulus case ($F(2, 28) = 4.35$, $p < .001$), but no interaction ($F(8, 112) < 1$). As expected, the optimal viewing position was slightly left of center (i.e., maximum performance is either on the second or third viewing position), and did not shift substantially as a function of stimulus case. The left panel of Figure 4 shows the stimulus case effects averaged over all viewing positions. Words in the IUC condition were recognized better than in the ALC and AUC conditions ($t(14) = 4.11$, $p < .001$, and $t(14) = 1.98$, $p < .05$, respectively, all t-tests one-tailed unless otherwise specified). However, the ALC and AUC conditions did not differ significantly

(t(14) < 1). In 14 out of 15 participants the, IUC condition yielded better accuracy than the average of the ALC and AUC conditions (see Figure 5).[9].

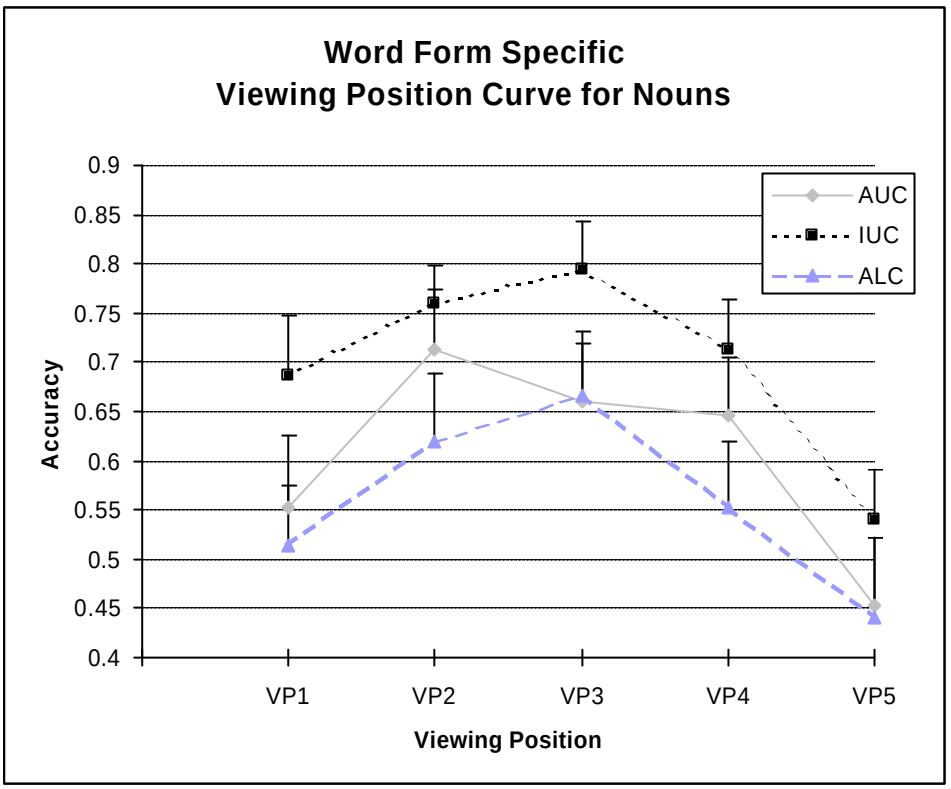


Figure 3: Word recognition accuracy for each word form as a function of viewing position with standard error margins. Words in the IUC condition are better recognized than both, words in ALC and AUC condition.

Because each stimulus was presented three times in three different cases, the possibility of repetition effects mediating the stimulus case effect must be examined. Therefore, I analyzed recognition of each stimulus case in an ANOVA together with their order of appearance in the experiment. There was a main effect of order of appearance ($F(2, 28) = 4.70$, $p = .02$) and stimulus case ($F(2, 28) = 10.71$, $p < .001$), but no interaction ($F(4, 56) = 1.31$, $p > .25$). Since each stimulus case had appeared approximately equally often in each rank of appearance ($Chi^2(8) = .22$) there was no hint in the data that the observed effects could be due to repetition of stimuli. Nevertheless, I performed an additional ANOVA for first appearances only to exclude any repetition influence: I obtained the same patterns of results for stimulus case as for all

[9] Note that item ANOVA does not make sense here, because due to the Latin Square design the prerequisite of independent, identically distributed (i. i. d.) random variables for items is not met, since different participants receive the same items under different viewing conditions. Therefore, item analyses are usually not computed in viewing position experiments (e.g., O'Regan & Jacobs, 1992). To perform an item analysis, items should be presented in identical conditions for each participant. Doing so in viewing position experiments would require to present each item in each viewing position for each participant. Then, repetition effects within one participant would make item analysis impossible, as repetition is again incompatible with the i. i. d. assumption. It is impossible to have both, a within-participant design where each participants sees all word forms (and identical words are repeated) and a design that allows item-based statistical analysis. To tackle this issue, Experiment 3 was designed in a way that allows individual item-based statistical analysis (see Experiment 3 and Appendix C).

stimuli. Again, the ANOVA revealed a main effect of stimulus case ($F(2, 28) = 6.41$, $p < .01$), driven by the superior performance of the IUC condition compared to the ALC and AUC conditions ($t(14) = 4.15$, $p < .001$, and $t(14) = 2.71$, $p < .001$, respectively). In contrast, the ALC and AUC conditions did not differ ($t(14) < 1$). Thus, none of these analyses indicated that the case effect was driven by stimulus repetition.

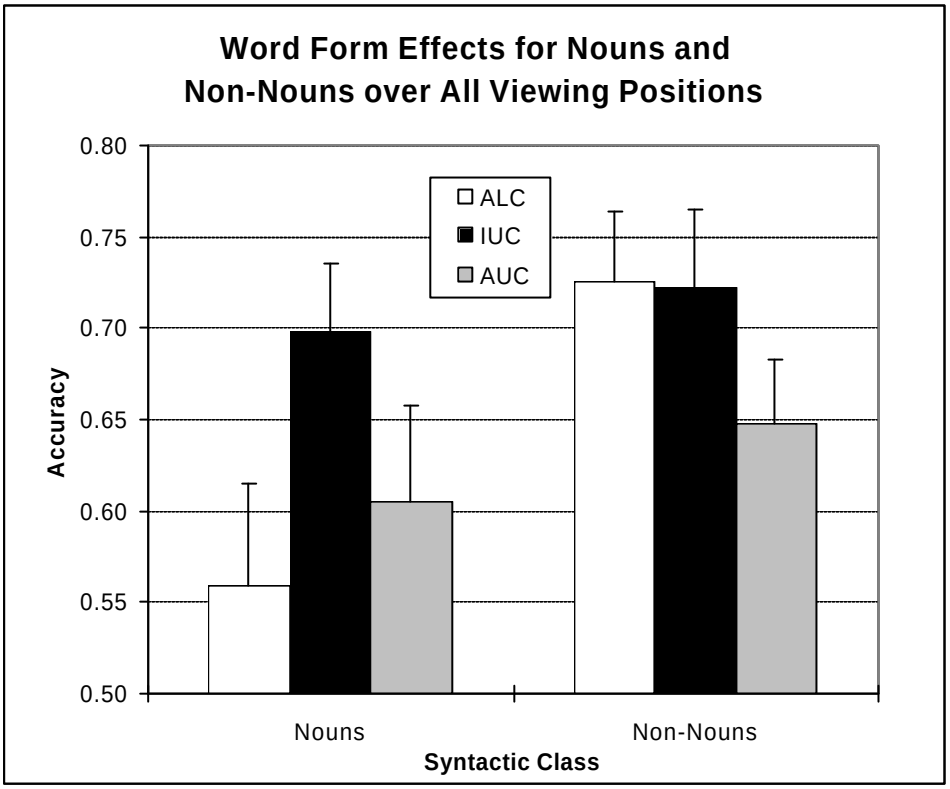


Figure 4: Mean recognition accuracy as a function of word form and syntactic class over all viewing positions with standard error margins. While the ALC condition clearly provides inferior performance for nouns (Experiment 1,) the performance for ALC non-nouns equals the performance in the IUC condition for non-nouns (Experiment 2).

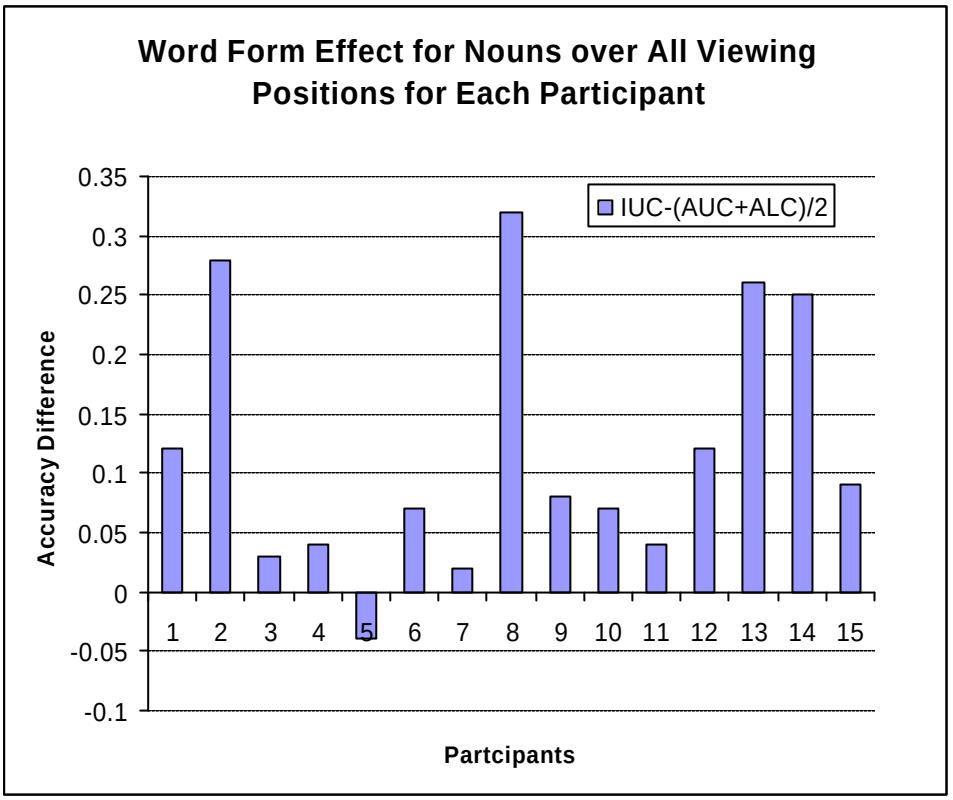


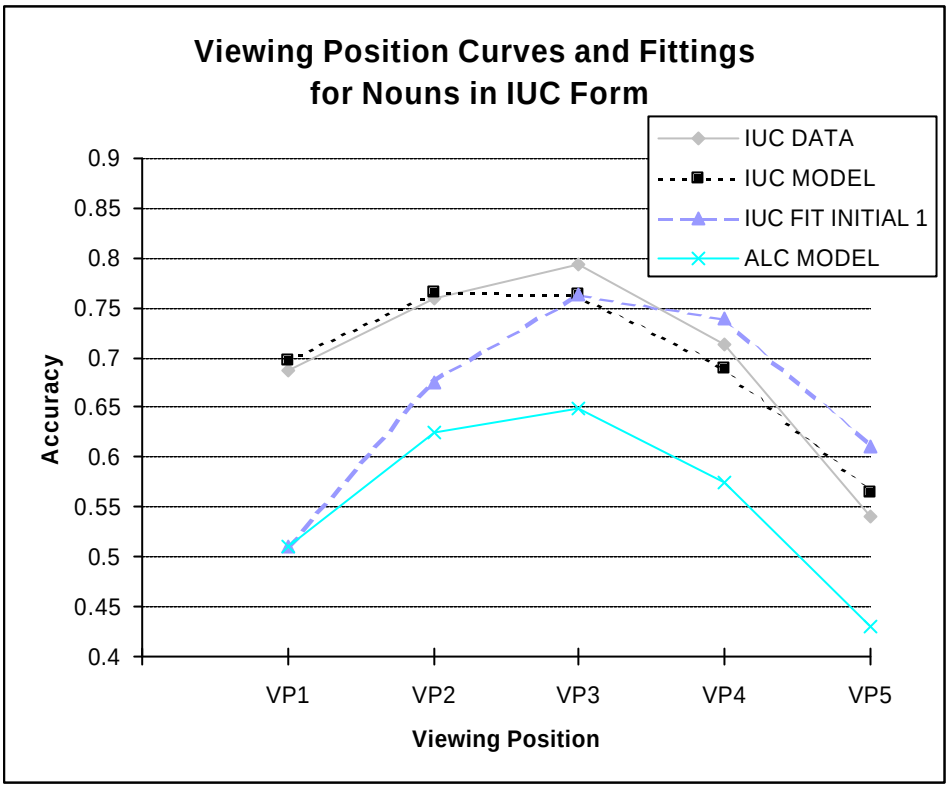
Figure 5: Word-form-specific effect for nouns for each participant over all viewing positions. The figure depicts the difference between the word form most frequently occurring in natural reading (IUC form) and the average of the forms almost never occurring in natural reading (ALC and AUC form).

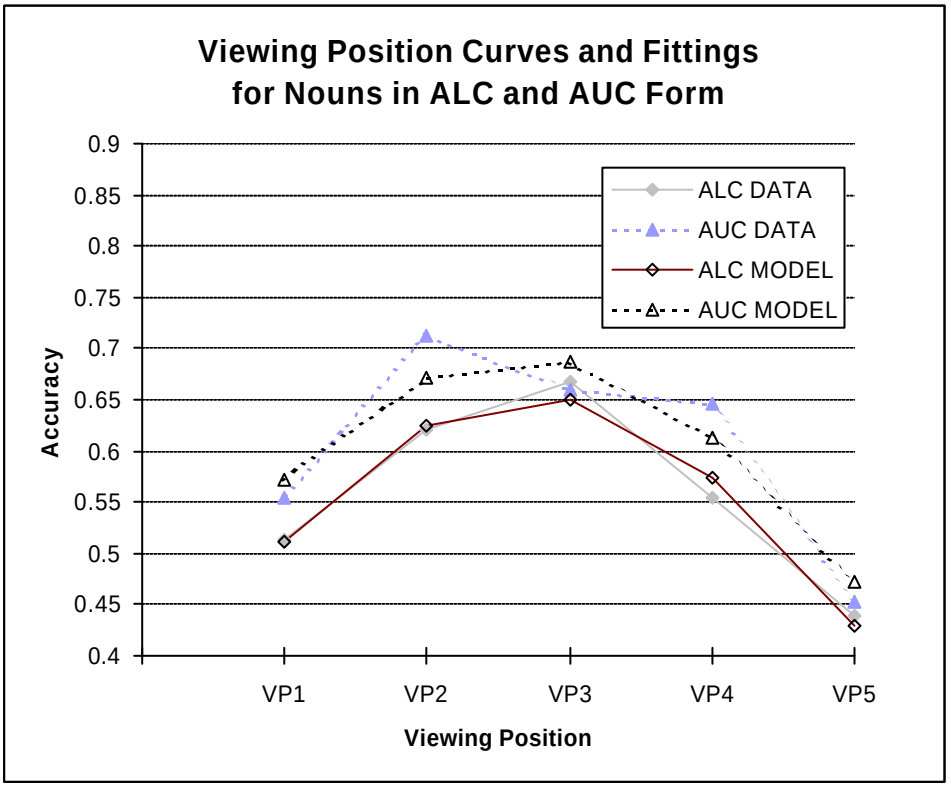
Since most PITs are vulnerable to sophisticated guessing that occurs post-perceptually (cf. Paap & Johansen, 1994), one could argue that this is not a perceptual difference, but rather that people accrue knowledge of the most common format of specific items and use it post-perceptually when they guess what they had seen. In particular, when a noun target is presented in ALC form they might be more likely to guess a non-noun neighbor of that target than when it is presented in the most common IUC form. Since most word recognition models (e.g., Grainger & Jacobs, 1996; Paap et al., 1982) essentially restrict wrong guesses to orthographic neighbors, an easy way to test this objection is then to look at the hermits in my stimulus group. As hermits have no neighbors, no form-specific sophisticated guessing effects, i.e., wrongly guessing neighbors in an uncommon presentation form, should be found for those subgroup of stimuli. However, if only the 12 hermits were analyzed, I found essentially the same results as for all stimuli despite the low power due to the small number of data points. Performance for the three stimulus cases differed ($F(2, 28) = 3.45$, $p < .05$) and again, nouns in the IUC condition were recognized better than in the ALC and AUC conditions ($t(14) = 2.25$, $p = .02$, and $t(14) = 2.27$, $p = .02$, respectively), while no difference occurred between ALC and AUC condition ($t(14) = 0.67$, $p = .27$). Thus, these results are not consistent with a sophisticated post-perceptual guessing account of the data.

2.4. Model Fits

Figure 6 and Table 2 show that the model fitted the ALC and AUC data well with low root mean square deviation (RMSD) values in both conditions. The model also fits the IUC data pattern well when assuming a lower drop-off rate parameter d than in the ALC and AUC conditions (see Figure 7 and Table 2). Most importantly, better recognition of the first letter alone as an explanation for the better performance in the IUC condition was not supported by the model fits. When recognition likelihood of the initial letter was fixed at 1 (IUC fit initial 1 condition; see Montant et al., 1998, for a similar argument) and the drop-off parameter of the ALC data was used for the remaining letters, the model predicted an asymmetrical curve with the optimal viewing position shifted to the right. Moreover, the asymmetry of the curve changed: Compared to the ALC curve, improvement in the IUC condition was highest on the final letter positions. Compared to the empirical data, however, the asymmetry of the fitted curve was shifted to the wrong side and yielded a poor fit (RMSD = 0.095). Overall performance was still estimated as too low by this model, although recognition probability of the capitalized first letter was set to the maximum (p = 1). If it was fixed to a lower value, inferior overall performance would even be worse. Particularly on the left fixation positions, this model could not fit the data.

Note that in the present fits the visual drop-off-parameter d was adjusted across the three conditions, although it should a priori be invariant because it is supposed to code pure letter legibility. However, that letter legibility dropped faster in the ALC condition (d = 0.061) than in the AUC condition (d = 0.052) is consistent with the basic ideas underlying the model because the drop-off rate is usually reduced when letters increase in size (Nazir et al., 1998). In contrast, that the drop-off rate fell further to a value of d = 0.034 in the IUC condition is not consistent with the basic ideas of the model because four of five letters have the same size in the ALC and IUC conditions and the drop-off rate for these letters should not differ. It is therefore concluded that the drop-off rate in the current form of the model of Nazir et al. (1998) is influenced by item-specific perceptual learning and not only by letter legibility itself.





Figures 6 (top) and 7 (bottom): *Word-form-specific viewing position curves for nouns showing word recognition accuracy for the different word forms as a function of viewing position together with model fits (see text for details).*

	Fitting Parameters for Nouns		
	d	A	RMSD
ALC	0.061	1.228	0.015
IUC	0.034	1.538	0.019
AUC	0.052	1.302	0.030
IUC - FIT - I1	0.067	1.919	0.096

Table 2: Best-fitting parameter values for nouns in the three form conditions of Experiment 1. d: Drop-off rate to the right; A: Asymmetry ratio; RMSD: Root mean square deviation; ALC: All lower-case condition; IUC: Initial upper-case condition; AUC: All upper-case condition. IUC – FIT – I1: Fit of the IUC condition with the parameters of the ALC condition and a recognition probability for the initial upper-case letter that is set to 1 (see text for details).

In sum, the observed effects could be accounted for by the model with varying drop-off rates for different word forms. The drop-off rate was not influenced by letter recognition alone. Optimal recognition of the initial letter alone in the IUC condition did not seem to suffice as an explanation for superior performance for two reasons. First, overall performance was still too low in the model if the ALC parameters were used and the recognition likelihood of the initial letter was lifted to 1. Second, the asymmetry of the viewing position curve should have changed markedly (to the right) if this assumption was made. However, there was no such hint in the data.

2.5. Discussion

The results of Experiment 1 and the model fits can be interpreted as follows. First, the present data establish an initial capitalization superiority effect and thus support the perceptual frequency hypothesis of item-specific word form: German nouns are best recognized in the perceptual form in which they are perceived most frequently, i.e., in IUC form. Second, there is no evidence that better legibility of the initial letter exclusively explains the performance in the IUC condition. Third, the fact that I failed to observe an advantage of the ALC condition over the AUC condition that is frequently observed in English (Paap et al., 1984) is also in line with the perceptual frequency hypothesis of item-specific word form. In contrast, general accounts neglecting item-specific word form would have made different predictions. If words were best perceived in same size, same case letters (or more specifically their resulting multiletter features), performance in the AUC and ALC conditions should not have been inferior to IUC presentation (Mayall et

al., 1997). Also, hypotheses about abstract single letter processing and greater distinctiveness of lower-case letters (Paap et al., 1984) fail to explain why ALC presentation yielded worse performance than IUC presentation.

Finally, the prototypical VPE (e.g., Nazir et al., 1998) showing an asymmetric curve was replicated for all word forms with German nouns. This is in line with the perceptual learning hypothesis of viewing location (Nazir et al., 1998; Nazir, 2000) because reading direction in German is also from left to right and, hence, the eye's landing distribution should be similar as in other Roman scripts, i.e., more towards the beginning of the word (Rayner, 1979). The fact that the prototypical viewing position curve was found even for unfamiliar word forms, supports Nazir et al.'s (1998, p. 820) claim that *„what is learned concerns the features of individual letters and where and how these features had been seen in the past."*

Therefore, the perceptual frequency hypothesis and the perceptual frequency hypothesis of visual word form must be distinct. That unfamiliar word forms produce a prototypical viewing position curve cannot be explained by perceptual learning of item-specific word forms. Vice versa, that different word forms produce performance differences in the way they did, cannot be accounted be an item-unspecific perceptual frequency hypothesis of location.

Together, the results of Experiment 1 suggest that features of individual letters as well as item-specific word form are perceptually learned and affect word processing. Thus, as discussed in the general introduction, perceptual learning can occur at various levels of the visual system (Ahissar & Hochstein, 1997). More specifically, perceptual learning in the reading process seems to be based on perceptually learned units coded at multiple levels (cf. Mayall et al., 1997). Nevertheless, there remain possible alternative accounts of my results so far: Language specificity and perceptual specificity. Finally, even though the model fits present evidence against a word initial letter legibility hypothesis, the latter cannot totally be excluded. These alternative accounts were tested in Experiment 2.

3. Experiment 2: Testing the Perceptual Frequency Hypothesis with German Non-Nouns in a PIT

3.1. Introduction

Experiment 2 was performed to test three alternative accounts of the data of Experiment 1. First, Experiment 2 tested a potential confound between language specificity and item-specific perceptual frequency. It is unclear, whether in German words are generally perceived better in IUC presentation or

whether this IUC superiority effect is restricted specifically to those words that are most frequently perceived in IUC form. Clearly, the perceptual frequency hypothesis of item-specific word form only holds if the latter prediction is true. Second, because performance is sensitive to the specific stimulus presentation conditions one can imagine that the results of Experiment 1 are specific to one or the other aspect of the display conditions rather than to word-specific visual form. If this objection is true, the same pattern of results should be observed for words that are less frequently encountered in IUC form. In contrast, if the perceptual frequency hypothesis is correct, IUC superiority should be restricted specifically to those words that are most frequently perceived in IUC form. Finally, as argued above, Experiment 2 should take away doubts as to whether superior performance in the IUC condition is not partially due to better legibility of initial upper-case letters.

The German language provides an easy way to test all three assumptions. Non-nouns (i.e., verbs, adjectives, adverbs) are commonly printed in ALC form. If they are used as a noun (e.g., “Das Schwimmen macht Spaß.“ <Swimming is fun>) or if they are the first word in a sentence they can also be printed in IUC form. Therefore, if the perceptual frequency hypothesis is correct performance for these non-nouns should be superior in ALC and IUC form compared to the less frequently observed AUC form, while ALC and IUC forms should not differ much. On the contrary, the same results as for nouns (i.e., IUC superiority only) should be observed according to language-specific, presentation-specific or word-initial bias accounts of the data of Experiment 1, because neither language nor presentation method was changed in Experiment 2.

3.2. Method

3.2.1. Participants

15 psychology students of the Philipps-University Marburg participated in Experiment 2. All were native German speakers and had normal or corrected to normal vision. They received course credit for participation. None of the participants had participated in Experiment 1.

3.2.2. Stimulus, Design and Procedure

50 German lemmas from the CELEX-Database (Baayen et al., 1993) with a mean log frequency per million of 1.76 were used in Experiment 2. Stimuli were either adjectives, verbs, or function words (see Appendix B). Design and Procedure were equal as in Experiment 1.

3.3. Results

Figures 4 (see above, right panel) and 8 show main effects of stimulus case ($F(2, 28) = 10.21$, $p < .001$) and viewing position ($F(4, 56) = 18.06$, $p < .001$), as well as an interaction ($F(8, 112) = 2.08$, $p < .05$). Figure 4 displays effects of stimulus case averaged over all viewing positions. Non-nouns were better recognized in the IUC and ALC conditions than in the AUC condition ($t(14) = 3.98$, $p < .001$, and $t(14) = 4.03$, $p < .001$). The ALC and IUC conditions, by contrast, did not differ significantly ($t(14) < 1$). Compared to Experiment 1 the major difference was found in the ALC condition. For nouns, lower-case presentation that almost never occurs in natural reading led to the worst performance. In contrast, for non-nouns, performance in the ALC condition was about as good as in the IUC condition. The difference between the ALC conditions in Experiments 1 and 2 can hardly be attributed to a general between-participant or a between-items difference because the two other conditions remained relatively stable (see Figure 8). As Figure 9 shows, in 13 of 15 participants the word forms frequently occurring in natural language (ALC and IUC) were better recognized than the less frequent AUC form.

Again, as expected, the optimal viewing position was slightly left from the middle, but this time it did shift as a function of stimulus case. Although the ALC and IUC conditions yielded better performance than the AUC condition for almost every data point (9 out of 10), this was particularly so for the initial two viewing positions (Figure 8: Differences ALC-AUC and IUC-AUC on initial vs. final two viewing positions, $t(14) = 2.19$, $p < .05$ and $t(14) = 2.59$, $p = .01$).

Given that each stimulus was again presented three times (once in each of the three form conditions), the possibility of repetition mediating the stimulus case effects was examined. The ANOVA revealed a main effect of order of appearance ($F(2, 28) = 23.38$, $p < .001$) and a main effect of stimulus case ($F(2, 28) = 15.03$, $p < .001$), but no interaction ($F(4, 56) < 1$). Since each stimulus case appeared approximately equally often in each rank of appearance ($Chi^2(8) = .88$) there was no evidence that the observed effects could be due to repetition. Nevertheless, I computed an additional ANOVA for first appearances only in which no repetition effects can be present. The results were as observed for all stimuli. A main effect of stimulus case ($F(2, 28) = 4.81$, $p = .016$) was due to better performance in the ALC and IUC conditions compared to the AUC condition ($t(14) = 3.27$, $p < .01$, and $t(14) = 2.04$, $p < .05$, respectively). The ALC and IUC conditions themselves did not differ ($t(14) < 1$). Thus, none of these analyses supported the hypothesis that stimulus repetition mediates the stimulus case effect for non-nouns.

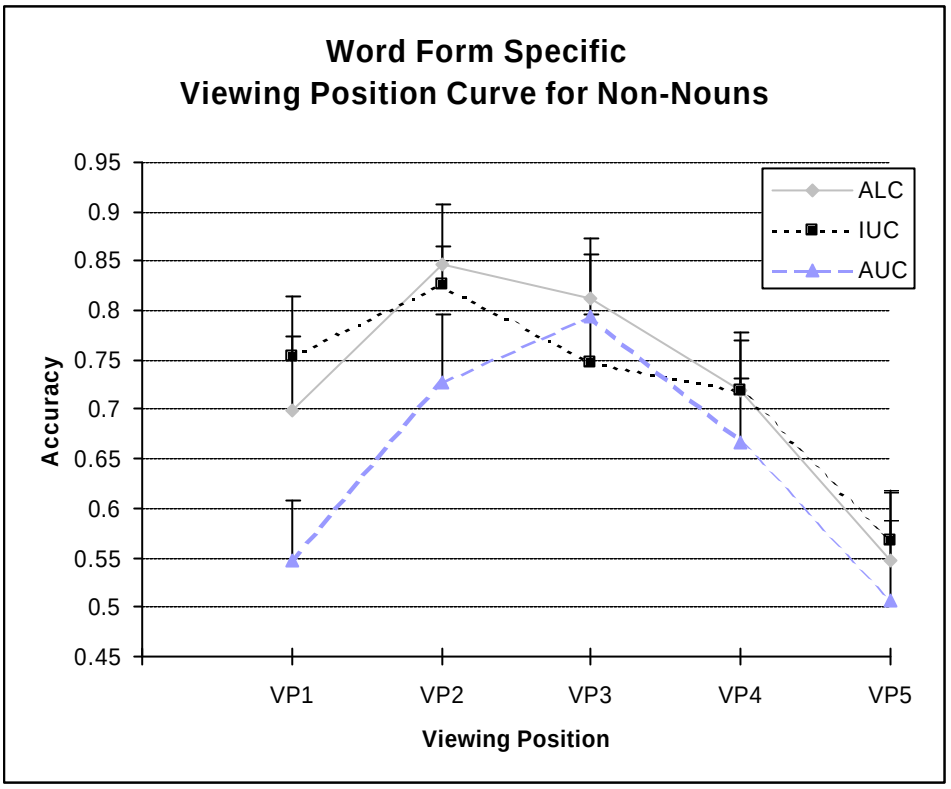


Figure 8: Word-form-specific viewing position curves for non-nouns showing word recognition accuracy for each word form as a function of viewing position with standard error margins.

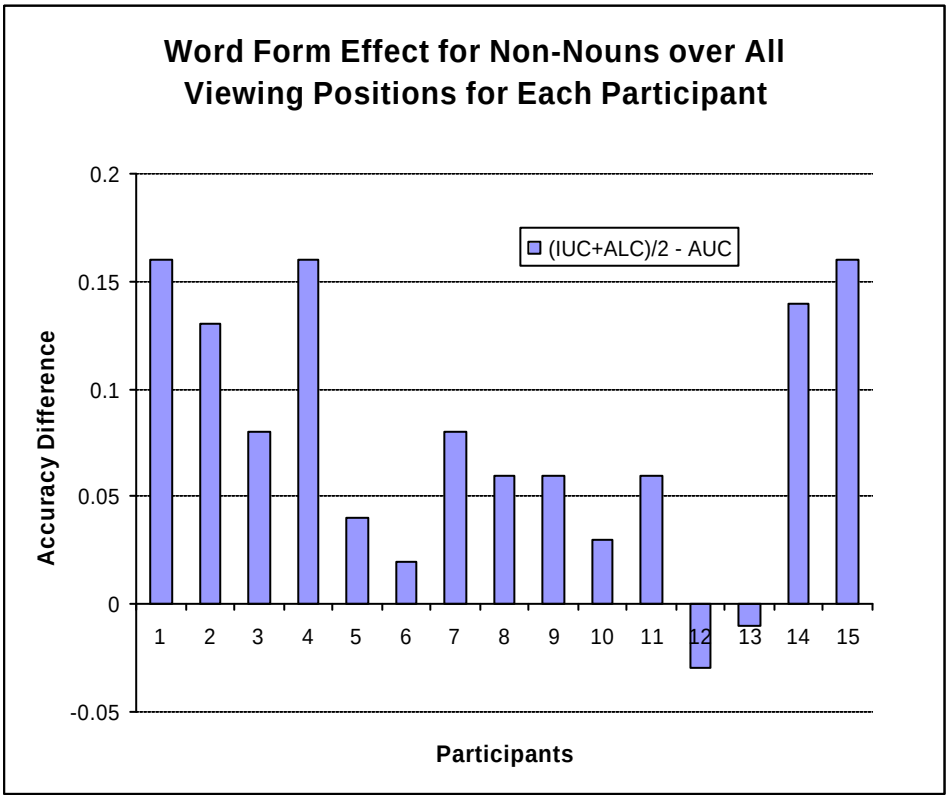
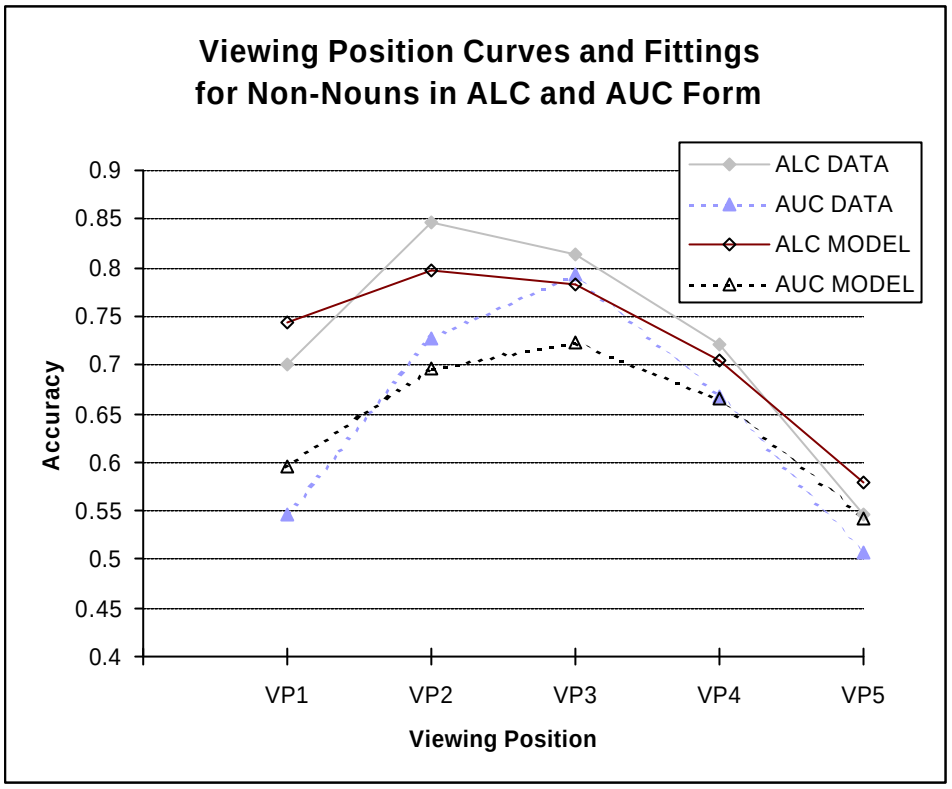


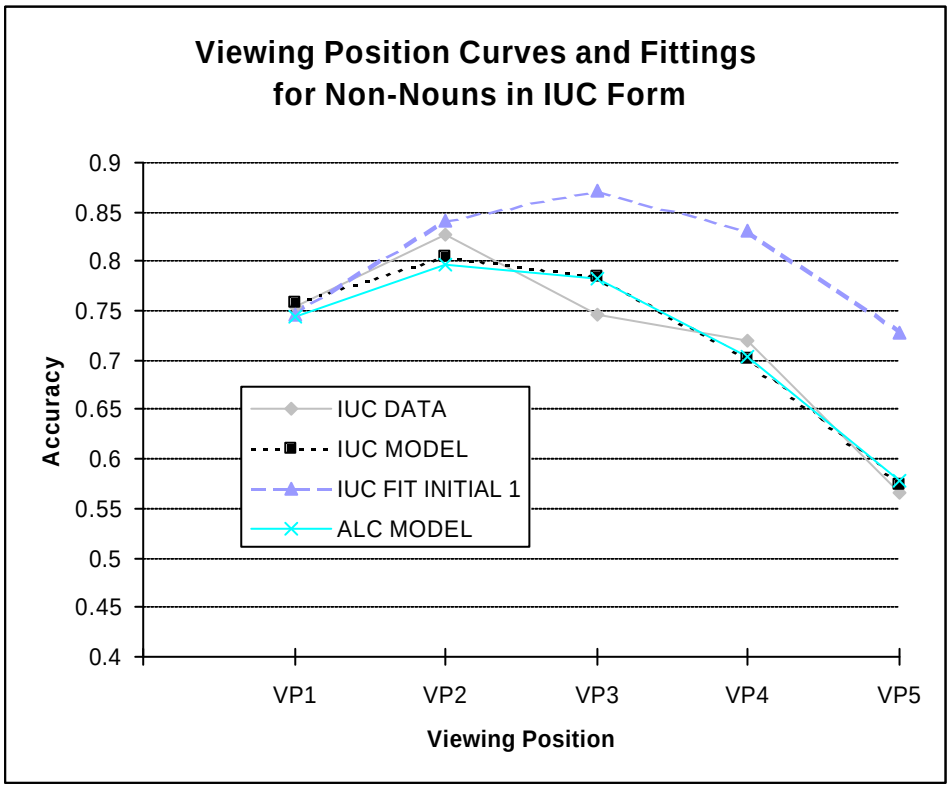
Figure 9: Word-form-specific effect for non-nouns for each participant over all viewing positions. The figure depicts the difference between the word forms most frequently occurring in natural reading (IUC and ALC form) and the form that almost never occurs in natural reading (AUC form).

To test a sophisticated post-perceptual guessing account of the data, I analyzed the subgroup of hermits in Experiment 2 (see result section of Experiment 1 for elaboration of this analysis). If only the 13 hermits were analyzed I found, in contrast to Experiment 1, no difference between stimulus cases ($F(2, 28) < 1$) and, in particular, again no difference between non-nouns presented in the IUC and the ALC condition ($t(14) = 0.53$, $p = .30$). The null effect in the ANOVA was due to the fact that for hermits the ALC condition did not differ from either IUC nor AUC condition (both $t(14) < .46$, both $p > .32$). While noun hermits in Experiment 1 produced better performance in IUC than in ALC form, non-noun-hermits did not show this pattern of results. Together, these results are not consistent with a sophisticated post-perceptual guessing account of the data.

3.4. Model Fits

The fits of the model of Nazir et al. (1998) to non-nouns were generally successful yielding RMSD values smaller than for the fits in Experiment 1 (see Table 3). The most important result was that the best-fitting parameter values for the ALC and IUC conditions reflected the empirical data: They did not differ much. Thus, performance in the IUC condition cannot be explained by the model with the assumption of superior recognition of the initial letter only because this assumption would lead to differently formed viewing position curves (see Figures 10 and 11 and Table 3). The model did again fairly well in fitting the empirical curves, although the fits were slightly worse than for nouns. For the ALC, IUC, and AUC condition the were fairly low (see Table 3). The drop-off rate for the AUC condition was substantially higher ($d = .048$) than for the ALC and IUC condition ($d = .028$ and $d = .027$), respectively. The drop-off rates did not differ much from Experiment 1 with the sole exception of the ALC condition in which it was much higher in Experiment 1 ($d = .061$). This difference cannot reflect different visual conditions since they were equal in Experiments 1 and 2. It can also not be simply explained by between-participant-differences or general between-item-differences (like imageability) since the drop-off rates in the other conditions in which items and participants were identical to the ALC condition, did not change. Thus, additional assumptions have to be made. One possible explanation for this result remains the perceptual frequency hypothesis of word-specific word-form perception: Because in the ALC presentation nouns have a perceptual frequency of 0 (and non-nouns have higher perceptual frequency) nouns may be recognized substantially worse when presented in that word form.





Figures 10 (top) and 11 (bottom): Word-form-specific viewing position curves for non-nouns showing word recognition accuracy for ALC, AUC and IUC word form as a function of viewing position and model fits. The model was able to fit the data pattern well. If recognition likelihood of the initial letter was fixed at 1 (IUC FIT INITIAL 1 condition) and if for the remaining lower-case letters drop-off parameters of the ALC data were used the fits did not differ much from the original curve, as ALC and IUC curve were empirically very similar.

	Fitting Parameters for Non-Nouns		
	d	A	RMSD
ALC	0.028	1.781	0.036
IUC	0.027	1.921	0.021
AUC	0.048	1.167	0.042
IUC - FIT - I1	0.028	1.781	0.100

Table 3: Best-fitting parameter values for non-nouns for the different conditions in Experiment 2. b_r: Drop-off rate to the right; A: Asymmetry ratio; RMSD: Root mean square deviation; ALC: All lower-case condition; IUC: Initial upper-case condition; AUC: All upper-case condition. IUC – FIT – I1: Fit of the IUC condition with the parameters of the ALC condition and a recognition probability for the initial upper-case letter that is set to 1 (see text for details).

Although all asymmetry ratios were between 1 and 2 reflecting the usual viewing position effect slightly left from the middle, the asymmetry ratios for the ALC and IUC conditions (A = 1.78, and A = 1.92) were somewhat higher than the asymmetry ratio for the AUC condition (A = 1.17). This difference in asymmetry ratio reflected nicely the difference in curve shapes (see Figure 8). While the AUC curve was almost symmetrical the ALC and IUC curves were much more asymmetrical: They differed from the AUC curve by a particularly good performance on the first two viewing positions. The difference between the AUC and the IUC condition could not be explained by better recognition of the initial letter because in that case particular improvement should have been observed on the final positions rather than the first position (see Experiment 1 for that argument in detail). Rather, the difference might be explained by an interaction of perceptual frequency hypotheses of location and word form: Words are much more often fixated on the initial fixation positions than on the final fixation positions (Nazir et al., 1998; Vitu et al., 1990). It is on those fixation positions where ALC and IUC word forms that occur frequently in regular reading improved most compared to the AUC word form.

3.5. Joint Analysis of Experiment 1 and Experiment 2

A joint analysis of Experiments 1 and 2 yielded significant effects of viewing position ($F(4, 112) = 35.30$, $p < .001$) and stimulus case ($F(2, 56) = 5.91$, $p < .001$), but no main between-participant effect of syntactic class ($F(1, 28) = 1.99$, $p > .1$). The interaction between syntactic class and case was also significant ($F(2, 56) = 4.44$, $p = .02$). It was mainly due to the difference between the syntactic classes in

the ALC condition ($t(28) = 2.42$, $p = .01$). In contrast, the AUC and IUC conditions did not differ between syntactic classes ($t(28) < .86$; $p > .40$ for both conditions). No other interaction reached significance.

The results of Experiment 2 had indicated an interaction between viewing position and stimulus case which had not been observed in Experiment 1. Performance in the ALC and IUC conditions was particularly better than in AUC condition on the initial viewing positions in Experiment 2, but not in Experiment 1 (see Figures 3 and 8). To further investigate this interaction, I directly compared the difference in performance in Experiments 1 and 2 for each stimulus case condition for those initial two viewing positions. T-tests showed that on the initial positions the ALC condition hinders performance for nouns in particular ($t(28) = 2.30$, $p = .01$, and $t(28) = 2.62$, $p < .01$, for position 1 and 2), while the IUC and the AUC conditions did not produce such an effect (all $t(28) < 1.14$, all $p > .10$).

Furthermore, I again explored the issue of repetition in the joint analysis. Just like in the separate analyses there were main effects of order of appearance ($F(2, 56) = 23.40$, $p < .001$) and stimulus case ($F(2, 56) = 12.92$, $p < .001$). The interaction of stimulus case and syntactic class ($F(2, 56) = 10.84$, $p < .001$) was also significant. The main effect of syntactic class ($F(1, 28) = 1.23$, $p > .10$) and all other interactions were not significant . Since each stimulus case form appeared approximately equally often on each rank of appearance ($Chi^2(8) = .88$) there was no hint in the data that the observed effects could be due to stimulus repetition. Still, I again computed an additional joint ANOVA for first appearances only and observed equal results as for all stimuli: A main effect of stimulus case ($F(2, 56) = 7.08$, $p < .01$) and an interaction of stimulus case with syntactic class ($F(2, 56) = 4.24$, $p < .05$). The examination of this interaction revealed that performance in the ALC condition improved for non-nouns ($t(28) = 1.64$, $p = .055$), while the AUC and IUC conditions did not differ between Experiments 1 and 2 (both $t(28) < 1$). Thus, much like the separate analyses, the results of the joint analyses allow to reject the hypothesis that stimulus repetition mediated the stimulus case by syntactic class interaction.

Finally, I again analyzed the responses to the 25 hermits which cannot be accounted for by (wrong) sophisticated post-perceptual guessing of noun and non-noun neighbors in inadequate perceptual forms. Despite the low number of data points an ANOVA revealed a tendency towards an interaction between syntactic class and word form approaches significance ($F(2, 56) = 2.42$, $p < .10$). This interaction was carried by the fact that the difference between IUC and ALC form was bigger for nouns than for non-nouns ($t(28) = 1.77$, $p < .05$). In sum, these results confirmed those of the single analyses that even in the absence of any form-specific sophisticated post-perceptual guessing of neighbors nouns were better

recognized in IUC than in ALC form while for non-nouns there was no such difference. Thus, again these results are not consistent with a sophisticated post-perceptual guessing account of the data.

3.6. Discussion

The results of Experiment 2 were strikingly different from those of Experiment 1. In Experiment 2, I observed that German non-nouns which are usually either printed in ALC or IUC form were better recognized in these more frequent word forms than in AUC form. Particularly, the difference in the ALC condition between Experiments 1 and 2 was striking: While performance for nouns that are almost never perceived in ALC form was worst in ALC presentation, performance for non-nouns that are frequently perceived in that ALC form was best (together with IUC form). Moreover, in contrast to Experiment 1, the results for German non-nouns replicated findings with English materials showing that lower-case presentation typically provides superior performance compared to upper-case presentation (Paap et al., 1984). In addition, the VPE was again replicated for all word forms with the typical left-right asymmetry. Finally, this time, I obtained an interaction between viewing position and word form: In the ALC and IUC conditions recognition particularly improved on the initial viewing positions compared to the AUC condition.

Thus, the data from Experiments 1 and 2 support the perceptual frequency hypothesis of item-specific word form: German words are best recognized in the perceptual form in which they are perceived most frequently, i.e., IUC presentation for nouns and ALC and IUC presentation for non-nouns. The perceptual frequency hypothesis also predicts the observed advantage of lower-case presentation over upper-case presentation only for those words that are frequently perceived in lower-case presentation, i.e., common English words (Paap et al., 1984) and German non-nouns. However, as predicted by the perceptual frequency hypothesis, no such advantage was obtained for words that are rarely perceived in lower-case presentation, i.e., German nouns. In contrast, any general account neglecting item-specific word form fails to provide different predictions for Experiments 1 and 2. If words were generally best perceived in same size and same case presentation (or more specifically with multiletter features of same size and same case letters), or if they were processed on the basis of abstract single letter units only, then the results of Experiments 1 and 2 should not differ in the way they did. Furthermore, if the lower-case advantage compared to upper-case presentation was only due to greater distinctiveness of lower-case letters (Paap et al., 1984), one should observe this advantage either in both or none of these experiments. Moreover, stimulus repetition cannot account for the word form effects, as an equal pattern

of results was obtained for the examination of first appearances only. Finally, language-specific, presentation-specific and word-initial bias accounts cannot explain the divergence between the results of Experiments 1 and 2 because this divergence was observed under equal presentation conditions within the same language.

Although the prototypical VPE showing an asymmetry was replicated for all word forms of German non-nouns, there seemed to be an interaction of viewing position with word form in Experiment 2: ALC and IUC forms were particularly better recognized on the initial positions. Because such an interaction was not found in Experiment 1, one needs to be careful with the interpretation. An interesting possible account of this finding postulates an interaction between perceptual frequencies of location and form of individual letter features within a particular word. Nazir et al. (1998) stated that *"what is learned comprises the features of individual letters and where and how these features had been seen in the past"* (p. 820). Fixations are much more often at the beginning of a word than at its end in Roman scripts (Rayner, 1979). Moreover, in German the individual letters and their features are commonly perceived in lower-case at the end of the word, while at the beginning of the word capitalization occurs inconsistently. Thus, I might have observed an interaction between „where" and „how" features have most often been seen in the past: What is most often seen are features of lower-case letters to the right of fixation. In contrast, features of upper-case letters are less often seen to the right of fixation. To the left of fixation both, features of lower-case and of upper-case letters, may be perceived. According to the perceptual learning hypothesis, one would expect superior performance at the location and in the form features are most often and most consistently seen. These features are those lower-case letters to the right of fixation, seen the ALC and IUC conditions with fixation on the initial letters. Given that such an interaction between viewing position and word form was not found in Experiment 1, this interpretation is tentative and needs further investigation.

In sum, the results of Experiments 1 and 2 support the perceptual frequency hypothesis of item-specific word form over other more general word form accounts because words were generally best perceived in the form in which they are most frequently encountered in natural reading. General accounts of visual word form perception fail to explain why the same word form manipulations have diverging effects on different words of equal lexical frequency.

4. Experiment 3: Testing the Perceptual Frequency Hypothesis of Item-Specific Word Form in a LDT

4.1. Introduction

As a further test of the perceptual frequency account, Experiment 3 was designed to generalize the results obtained in Experiments 1 and 2 in two respects: First, following the strategy of functional overlap modeling (Grainger & Jacobs, 1996; Jacobs & Grainger, 1994), I wanted to replicate the findings from the two PITs in another task (the LDT) in order to demonstrate task-overlapping rather than task-specific effects. Because each task affords both task-specific and task-independent processes only a multitask approach might enable me to detect task-independent processes generalizable to normal reading. In order to constrain models of word recognition, it is important to be able to interpret the word form effects found in Experiments 1 and 2 as general rather than task-specific visual word recognition effects. As discussed in the general introduction, case-mixing effects can differ between different tasks (for a review, see Mayall & Humphreys, 1996a) and such differences have motivated hypotheses about additional pathways working in some tasks but not others (e.g., Besner & Johnston 1989; Besner & McCann, 1987, for the familiarity checking mechanism).

Second, to exclude the possibility of a presentation-specific participants' strategy I presented nouns and non-nouns in mixed rather than blocked mode. Strategic influences on performance have received much attention in recent years (Gibbs & Van Orden, 1998, for a review). Even when response modes and stimuli are held constant, task demands can influence performance. A prominent example are responses in the LDT depending on the quality of nonwords (e.g., legal vs. illegal vs. pseudohomophones) that alters frequency effects (e.g., Stone & Van Orden, 1993), neighborhood effects (e.g., Andrews, 1997; Grainger & Jacobs, 1996), and regularity effects (e.g., Gibbs & Van Orden, 1998). Concerning word form effects, participants also seem to be sensitive to strategic manipulations, as suggested by the diverging results of Kinoshita (1987) and Besner and McCann (1987). Therefore, I wondered whether specific task demands and stimulus selection or an interaction between them were responsible for the observed effects in Experiments 1 and 2. Particularly, Experiment 1 used a blocked design with nouns which are most frequently perceived in IUC form. If participants had become biased towards the most frequent perceptual form of nouns (IUC form) in that experiment, the results would reflect a task-specific rather than an item-specific word form effect. Similar arguments could be made for the blocked non-noun presentation. This

possible problem can be tackled with a mixed design in which nouns and non-nouns are presented in pseudorandom order.

Moreover, I wanted to address two further possible methodological objections that could be raised against the Experiments 1 and 2. First, although no hint of repetition effects interacting with word form effects was observed, one could wonder whether there is some non-linear influence of repetition on word form that is not detected in an ANOVA based on the general linear model. Second, because of the main repetition effect and viewing position alternations within and between participants for each item, item analysis for different word forms did not make sense (see Footnote 3). Such an analysis might nevertheless provide crucial information. Therefore, in Experiment 3 the within-participant design for different word forms was replaced by a design allowing to directly compare results for different word forms on an item-based level.

Finally, Experiment 3 explored two conceptual questions not addressed in Experiments 1 and 2: i) what is the interaction between lexical frequency and perceptual frequency, and ii) how specific is item-specific perceptual frequency, i.e., does it generalize to similar nonwords. Concerning the first question, in the introduction I argued that perceptual frequency and lexical frequency may index two opposite interactive processes which may cancel each other out when analyzed in an ANOVA using the general linear model. Experiment 3 empirically investigated this objection: The argument that a null interaction between a perceptual manipulation (e.g., case mixing) and a lexical manipulation (e.g., frequency) excludes an item-specific perceptual word form effect would be difficult to defend if I obtained such a null interaction while still replicating the main item-specific word form effects of Experiments 1 and 2. Concerning the second question, the results of previous research summarized in the introduction indicated that specificity of perceptual learning varies. For the nonwords, I used nonword neighbors of nouns and non-nouns with a particular item-specific word form to investigate the degree of specificity. Thus, I examined whether my case manipulations only and specifically affected words frequently read in that word form themselves, or whether they also generalized to similar nonword forms.

4.2. Method

4.2.1. Participants

16 psychology students of the Philipps-University Marburg participated in the Experiment. All were native German speakers and had normal or corrected to normal vision. They received course credit for participation. None of them had participated in Experiments 1 and 2.

4.2.2. Stimuli and Design

80 German five-letter nouns and 80 five-letter non-nouns were used in the experiment. Frequency and the reference lists for computation of all other measures were taken from the CELEX-Database (Baayen et al., 1993). As concerns words, frequency and syntactic class (nouns vs. non-nouns) were manipulated in a 2*2 design, leading to four groups of 40 words each (Table 4, Appendix B1, B2). Low-frequency item groups included only words with 10 or less occurrences per million while high-frequency item groups consisted of words with more than 50 occurrences per million. Number of neighbors, number of phonemes, and number of syllables were matched between all stimulus groups, and log frequency was manipulated between the respective stimulus groups. Furthermore, the number of higher frequent neighbors was matched between the respective frequency groups. 20 of those 40 words of each group were presented in IUC form and the other 20 in the ALC form in stimulus list 1, and vice versa in stimulus list 2. The stimulus lists were varied between participants (Latin Square Design), so that each word was presented equally often in ALC and IUC form across participants. This procedure enabled me to test the hypotheses of word-specific form effects directly in a within-item analysis for each stimulus group by using the Loftus and Masson method (Loftus & Masson, 1994).

Nonwords were created by exchanging one letter from each target stimulus at Positions 2, 3, 4, or 5. For each stimulus group of 40 words 10 nonwords were created by exchanging a letter at one of those four positions. Number of neighbors and number of syllables were matched between nonword groups. The number of neighbors was held as low as possible to create a maximum orthographic similarity between the nonwords and words they were derived from and a maximum orthographic dissimilarity to all other words (see Table 5, Appendix B3, B4). The variation of ALC and IUC form was the same as described above for words.

WORDS		Nouns		Non-Nouns	
		high F	lowF	high F	low F
logF	Means	2.06	0.55	2.08	0.55
	StdErr	0.05	0.05	0.04	0.05
N	Means	1.93	1.93	1.85	1.85
	StdErr	0.35	0.19	0.36	0.38
HFN	Means	0.25	0.95	0.33	0.98
	StdErr	0.09	0.13	0.10	0.30
Phonemes	Means	4.55	4.65	4.58	4.65
	StdErr	0.08	0.09	0.09	0.08
Syllables	Means	1.73	1.80	1.73	1.73
	StdErr	0.07	0.06	0.07	0.07

Table 4: Means and standard errors of word stimuli used in Experiment 3. High F: High-frequent words, low F: Low-frequent words, logF: Logarithmic word frequency, N: Number of orthographic neighbors, HFN: Number of higher frequent orthographic neighbors; Phonemes: Number of phonemes; Syllables: Number of syllables.

Nonwords		Nouns		Non-Nouns	
		high F	lowF	high F	low F
Base logF	Means	2.08	0.55	2.06	0.55
	StdErr	0.04	0.04	0.05	0.05
N	Means	1.63	1.58	1.75	1.65
	StdErr	0.13	0.14	0.15	0.13
Position	Means	3.50	3.50	3.50	3.50
	StdErr	0.18	0.18	0.18	0.18
Phonemes	Means	4.58	4.68	4.55	4.65
	StdErr	0.09	0.07	0.08	0.09
Syllables	Means	1.73	1.73	1.73	1.80
	StdErr	0.07	0.07	0.07	0.06

Table 5: Attributes of nonword stimuli used in Experiment 3. High F: High-frequent words, Low F: Low-frequent words, logF: Logarithmic word frequency, N: Number of orthographic neighbors; Position: Letter position at which a letter was exchanged from a baseword. Note that, since each nonword was created by exchanging one letter from a baseword, each nonword has at least one orthographic neighbor. While it was not always possible to find nonwords with the exchange letter at the adequate position that have no other neighbor than the baseword, neighborhood was held as low as possible.

4.2.3. Procedure

The experiment was run on an Apple Power PC 7200/90. Participants sat approximately 30 cm from the screen. A trial began with a fixation colon in the middle of the screen. After 500 ms the colon disappeared and a stimulus was presented in Courier 24 font on central viewing position. The stimulus remained visible until participants pressed one of two keys indicating whether the stimulus was a word or a nonword. After an interstimulus interval of 500 ms the next trial began. Participants received 16 training trials prior to the randomized 320 experimental trials. Participants were instructed to respond as quickly and accurately as possible. RTs slower than 2000 ms were not recorded. The whole experiment took about 20 minutes.

4.3. Results

Items with an overall accuracy below .67 were eliminated, i.e., responses to each item had to be correct at least twice as often as incorrect. This cleaning procedure affected only four low-frequency non-nouns, but no nouns, no high-frequency non-nouns, and no nonwords. On the basis of this item pool mean correct RTs were computed. The trimming procedure excluded scores smaller than 200 ms and greater than 3 SDs above the participant's overall mean. ANOVA over participants and one-sided t-tests over items and participants were separately conducted for words and nonwords and nouns and non-nouns.

4.3.1. Words

An ANOVA over participants for all words indicated main effects of frequency ($F_1(1, 15) = 85.49$, $p < .001$) and syntactic class ($F_1(1, 15) = 55.91$, $p = .001$), but not of stimulus case ($F_1(1, 15) = 2.46$, $p = .14$). High-frequent words were responded to faster than low-frequent words, and nouns produced faster RTs than non-nouns. Most importantly, I observed a significant interaction between case and stimulus type ($F_1(1, 15) = 20.39$, $p < .001$). Nouns produced faster RTs in the IUC condition, while non-nouns produced faster or at least equally fast RTs in the ALC condition (see Figure 12). No other interaction reached significance (all $p > .15$). A separate ANOVA for nouns yielded main effects of frequency and case ($F_1(1, 15) = 75.12$, $p < .001$, $F_1(1, 15) = 85.49$, $p < .001$, respectively). A separate ANOVA for non-nouns yielded a main effect of frequency and a marginally significant effect of case in the opposite direction as for nouns ($F_1(1, 15) = 53.12$, $p < .001$, $F_1(1, 15) = 3.25$, $p = .09$), respectively. In both separate analyses, I observed no interaction between frequency and case (both $p > .15$).

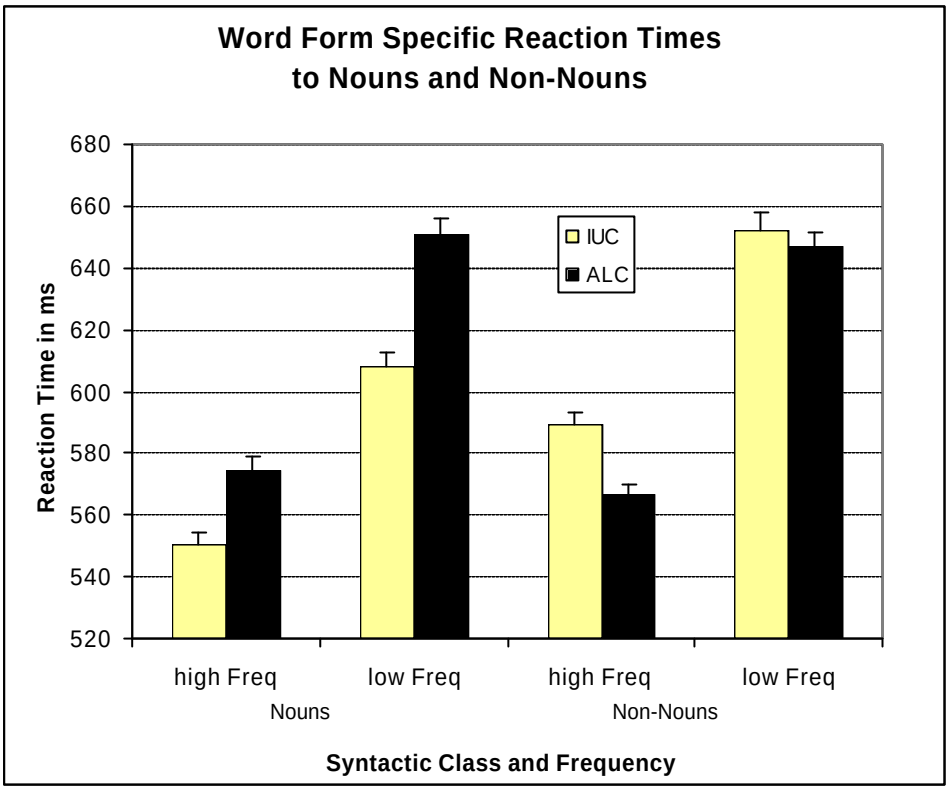


Figure 12: Lexical decision times (RTs) for nouns and non-nouns of different frequencies.. For nouns IUC form produced faster responses than ALC form, while for non-nouns no such difference (low-frequency group) or even the reversed pattern could be observed (high-frequency group).

To directly test the hypothesis that word form affects nouns and non-nouns differentially I computed t-tests within participants (t_1) and within items (t_2). For the within-item analysis I normalized RTs (Loftus & Masson, 1994). Basically, the method normalizes the item RTs for each participant by shifting them to the same grand mean. Any difference in the within-item analysis can then no longer be attributed to differences between participants that received a particular item in one or the other case condition.

The IUC condition produced better performance than the ALC condition for nouns in both analyses, for both the high-frequency items ($t_1(15) = 4.02$, $p < .001$; $t_2(39) = 3.27$, $p = .001$), and the low-frequency items ($t_1(15) = 3.14$, $p < .01$; $t_2(39) = 1.48$, $p < .001$). On the contrary, for non-nouns, the ALC condition produced faster RTs than the IUC condition for high-frequency non-nouns ($t_1(15) = 2.27$, $p < .05$; $t_2(39) = 3.09$, $p < .01$), but not for low-frequency ones ($t_1(15) < 1$; $t_2(39) < 1$). The difference between the IUC and the ALC conditions for nouns and non-nouns is shown in Figure 13. While 12 of 16 participants produced an IUC superiority effect for nouns, for non-nouns the effects were more variable (see Figure 13).

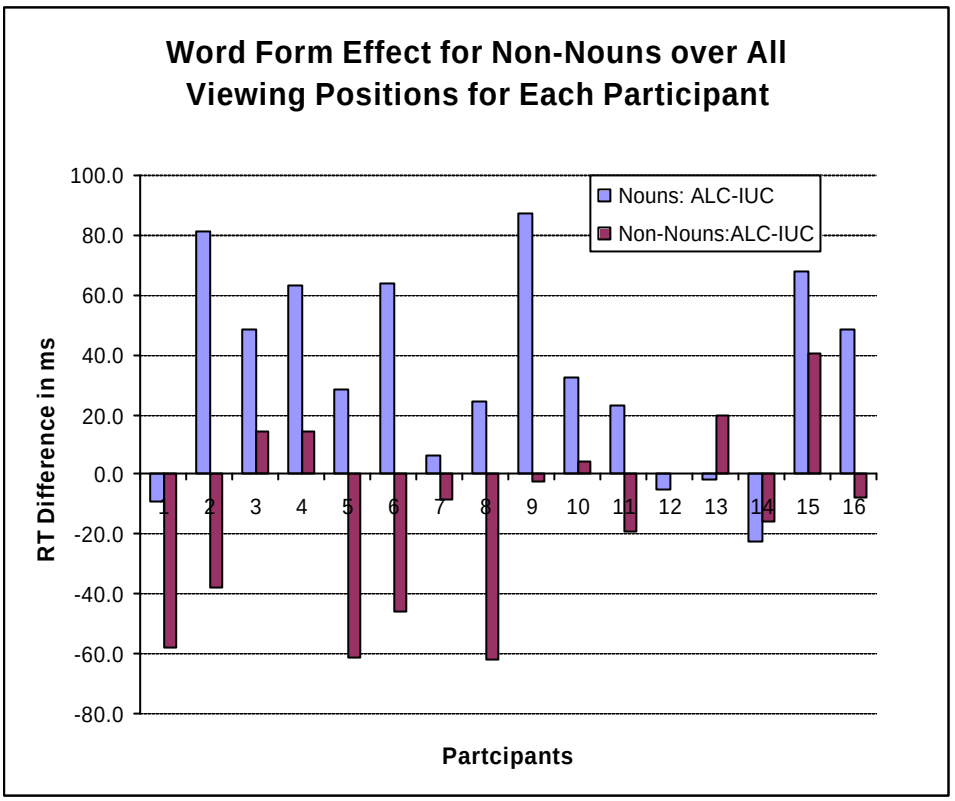


Figure 13: Word-form-specific effect for nouns and non-nouns for each participant over all viewing positions. The figure depicts the difference in RT between the IUC word form most frequently occurring in natural reading and the ALC form almost never occurring in nouns. The same difference is computed for non-nouns where both forms can occur.

The error analyses did not provide any evidence for a stimulus-dependent speed-accuracy trade-off. If there were any effects of stimulus case, they pointed into the right direction: There was a marginally significant effect of case in the separate ANOVA for nouns ($F_1(1, 15) = 3.08$, $p < .10$) that was carried by a superior accuracy for low-frequent nouns in the IUC condition ($t_1(15) = 1.96$, $p < .05$; $t_2(39) = 1.70$, $p < .05$). All other case effects were not significant.

Altogether, the results were straightforward: Case effects interacted with syntactic class in a similar way as in Experiments 1 and 2. Nouns were best recognized in IUC form, while non-nouns were equally well or better recognized in ALC form. T-tests indicated that high-frequent non-nouns were even recognized better in ALC presentation than in IUC presentation, while for low-frequent non-nouns the case conditions did not make much difference.

4.3.2. Nonwords

An ANOVA for all nonwords indicated no significant effects of baseword frequency, baseword syntactic class, or baseword stimulus case, and no interaction of basewords stimulus case and syntactic

class (all $F_1(1, 15) < 2.06$, all $p > .17$, see Figure 14). Separate ANOVAs for baseword nouns and baseword non-nouns revealed no significant effects of case or frequency (all $F_{1s}(1, 15) < 1.77$, all ps > .20), except for a marginally significant frequency effect for baseword nouns: Nonwords derived from high-frequent nouns seemed to be harder to reject than nonwords derived from low-frequent nouns ($F(1, 15) = 4.21$, $p = .065$). T-tests within participants (t_1) and items (t_2) using the Loftus-Masson method (Loftus & Masson, 1994) generally confirmed those results: No indication of an effect of stimulus case was observed for any stimulus group (all $t_1(15) < 1.22$, all $p > .24$; all $t_2(39) < 1$), except for a marginally significant effect for nonwords derived from high-frequent baseword nouns in the item analysis ($t_2(39) = 1.54$, $p = .066$). In the error analyses, there was also no indication of a significant baseword form effect.

Mayall et al. (1997) recently hypothesized that word form effects are due to a disruption of transletter features. If this is correct, one could argue that the main feature of the word form is not really the initial letter only, but rather the transletter feature(s) it constitutes together with its neighbor letters on second or third letter position. However, exchanging one of the final two letters (fourth or fifth position) of a word to create a neighbor nonword should not disrupt this distinctive transletter feature(s). Therefore, for those nonwords, I reanalyzed all effects. However, neither the results of the ANOVA nor the results of the t-tests changed substantially: No significant effect of stimulus case on RTs or errors was observed for any stimulus group (all $F_1(1, 15) < 1.54$, all $p > .23$). Separate ANOVAs for baseword nouns and baseword non-nouns also revealed no significant effect of case and frequency (all $F_1(1, 15) < 2.04$, all $p > .17$). T-tests within participants (t_1) and within items (t_2) using the Loftus-Masson method (Loftus & Masson, 1994) confirmed those results: No significant effect of stimulus case was observed for the specific tests for any stimulus group (all $t_1(15) < 1$; $t_2(19) < 1.09$, all $p > .29$. Again, there were also no effects in the error analysis.

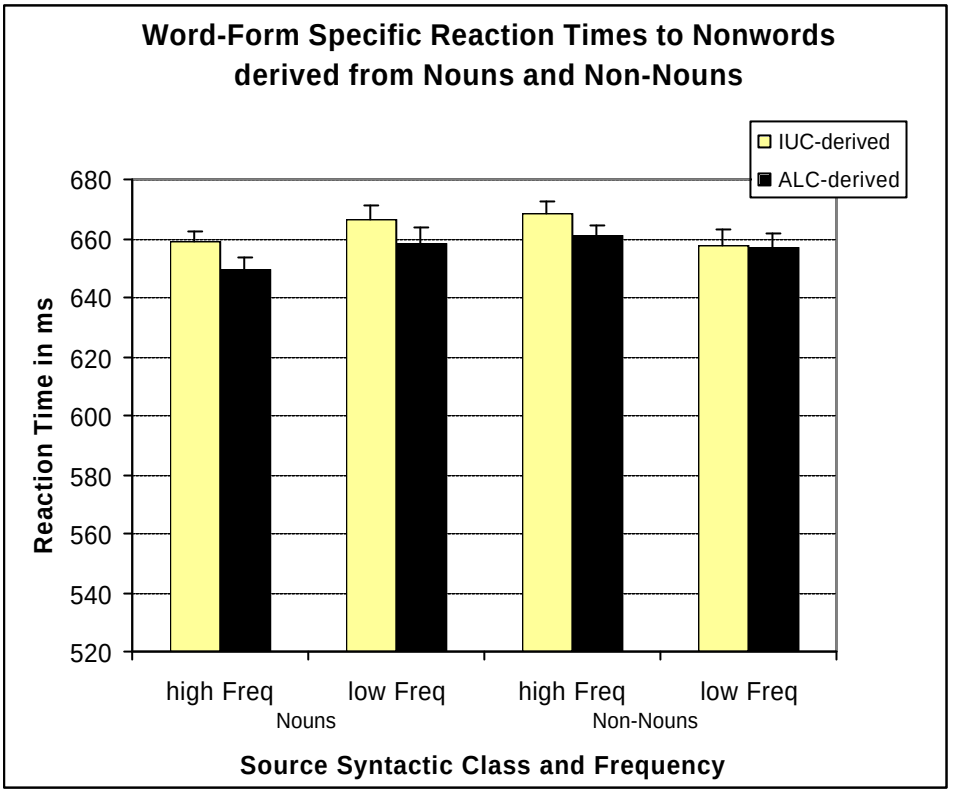


Figure 14: Lexical decision times (RTs) for nonwords derived from nouns and non-nouns of different frequencies.

Thus, case did not seem to have an influence on rejection times or errors of highly similar nonwords in any analysis, except for the high-frequency noun baseword condition in the item analysis. However, because this effect was not significant, the present results provide no evidence for a transfer of item-specific word form effects to similar nonwords.

4.4. Discussion

Experiment 3 yielded four important results: First, nouns were recognized faster in IUC presentation than in the ALC presentation, in both, the participant and item analysis. Hence, German nouns were again not perceived best in same size, same case presentation. Second, non-nouns were not recognized faster in IUC presentation than in ALC presentation. On the contrary, ALC presentation provided superior performance for high-frequency non-nouns. The observed difference between nouns and non-nouns is consistent with the perceptual frequency hypothesis, but inconsistent with any account focussing only on general aspects of word form perception. Third, the interaction between frequency and case manipulation in the ANOVA did not reach significance. Therefore, using the additive factors logic critically discussed in the introduction, one would conclude that there are no item-specific word form effects . However, the presence of item-specific effects for the different syntactic classes clearly is in contradiction with this

conclusion. Fourth, no significant effect of case manipulation was observed for nonwords, suggesting that the perceptual frequency of word form is really item-specific and does not transfer to similar nonwords.

5. General Discussion of Study 1

Two PITs and one LDT provided evidence for the perceptual frequency hypothesis of item-specific visual form: The more often a word was previously perceived in a particular item-specific visual word form the better it was recognized in that particular form. German nouns which are usually encountered with capitalization of the initial letter were best recognized in that particular visual form both, a PIT and a LDT. In the PIT, IUC form provided better performance than both AUC and ALC form and in the LDT IUC form was superior to ALC form. In sharp contrast to German nouns, a very different pattern of results was obtained for German non-nouns which can be encountered either in ALC form or in IUC form. Non-nouns are written in IUC form if they are functionally used as a noun ("Das Schwimmen macht Spaß." <Swimming is fun>) or if they are at the beginning of a sentence. For non-nouns, IUC and ALC form did not differ in the PIT, and in the LDT for high-frequency non-nouns the pattern was even reversed: ALC form led to faster responses than IUC form for high-frequency non-nouns. Also, the comparison between all lower-case and all upper-case presentation differed between syntactic classes in the PIT. Non-nouns were like common English words (Paap et al., 1984) better recognized in lower-case than in upper-case, while no such difference was found for nouns – in contrast, descriptively the mean difference even was in the opposite direction. In sum, the results are consistent with the perceptual frequency hypothesis of item-specific word form: Nouns and non-nouns are differentially better recognized in that particular word form they are encountered most often.

Thus, the results favor a perceptual frequency hypothesis of item-specific visual word form. General accounts of case mixing or perceptual distortion fail to explain the difference between nouns and non-nouns found in Study 1. Altogether, Study 1 implies that the visual variable perceptual frequency may play an important role in visual word recognition. The methods of its theoretical and empirical investigation are strongly questioned by this study because the interpretation of null interactions with frequency using AFL were not appropriate in Study 3. The consequences for the theoretical assumptions and models underlying the investigations of visual determinants will be discussed in detail in the general discussion. First, however, it will be investigated whether visual determinants do also modulate standard hallmark effects of visual word recognition: If this was the case, such findings would again question the regular applicability of the general linear model and AFL in this research domain.

III. Study 2: The Alteration of the Frequency Effect and its Interactions with Visual Variables

1. Introduction

In the general introduction, I have reviewed the literature and argued that frequency in most tasks particularly exerts its top-down effects under non-optimal conditions. As I am mainly concerned with the aspects of visual alteration, I will test this hypothesis for non-optimal visual presentation conditions, namely different viewing positions. Stronger frequency effects under poor presentation conditions would be in line with connectionist model simulations (Montant et al., 1998) and, qualitatively, also with the model of Nazir et al. (1998, see introduction for details). However, it is not clear whether the model also captures the alterations of the frequency effect by viewing position quantitatively well. Therefore, viewing position data and fits for stimulus groups with different frequencies will be provided in Study 2.

Furthermore, null interactions of frequency with other variables have been interpreted using AFL to conclude that lexical processes do not interact with other processes at the same stages. In the introduction, I have argued that null interactions are particularly likely to occur under near-optimal presentation conditions, too. Therefore, interactions may be missed in optimal viewing positions. This hypothesis will also be investigated in Study 2 by examining the interaction of frequency and word length at different viewing positions. If visual factors, such as viewing position, do not influence frequency effects and their interactions, we should not observe any differences between favorable and less favorable viewing positions. However, if the interaction of the frequency effect with other variables varies with viewing position, studies using only fixation on the same, central, near-optimal viewing position may underestimate the interaction between lexical factors and others.

2. Experiment 4: Testing the Alteration of the Frequency Effect and its Interaction with Word Length on Different Viewing Positions

2.1. Introduction

Experiment 4 was designed to study these hypotheses: Firstly, to my knowledge, the modulation of interaction of frequency with word length as a function of viewing position has not yet been investigated in a PIT. Based on the review of frequency effects, word length effects, their interaction effects with other

variables and of how these effects are altered by viewing position, the predictions are straightforward: All three effects should increase with poorer viewing position. More specifically, it will be examined whether these effects are also obtained on central viewing positions or only on other viewing positions. Secondly, in the mathematical model for VPEs in PITs, no invariant frequency parameter has been incorporated yet (Nazir et al., 1998, see introduction). In the introduction, I generalized the model to any combination of viewing position and word length, so that it could be applied to different word lengths. Investigating the specific parametrical nature of the frequency effect as a function of viewing position and word length will form restrictive conditions for a future mathematical viewing position model, whose scope is extended to also account for frequency effects. Finally, neighborhood size (cf. Coltheart et al., 1977) is naturally negatively correlated with word length, but often not controlled when both frequency and word length are investigated. Therefore, I controlled neighborhood size to ensure that any interaction with word length is not driven by a neighborhood × frequency interaction (cf. Grainger et al., 1989, 1992).

2.2. Method

2.2.1. Participants

15 psychology students of the Philipps-University Marburg participated in the experiment. All were native German speakers and had normal or corrected to normal vision. They received course credit for participation in the experiment.

2.2.2. Stimuli and Design

150 German lemmas from the CELEX- Database (Baayen, et al., 1993) were used in Experiment 4. The stimuli were three-, five- and seven-letter German words, manipulated for frequency and matched for orthographic neighborhood (see Table 6 and Appendix D). Each group in a 2 × 3 ANOVA design (three word length by two frequency groups) consisted of 25 stimuli.

The viewing position alteration using a Latin Square Design for five sublists with five stimuli from each stimulus group was identical to Experiments 1 and 2: The words were subdivided in five equally spaced areas, and the five viewing positions were at the center of those areas (cf. O'Regan & Jacobs, 1992).

	Low Frequency			***High Frequency***		
	Number of Letters			*Number of Letters*		
	3	5	7	3	5	7
log F	0.18	0.18	0.18	2.76	2.86	2.50
N	2.76	2.80	2.72	2.72	2.72	2.72
N-Grad	-0.16	-0.27	-0.05	-0.28	-0.05	-0.06

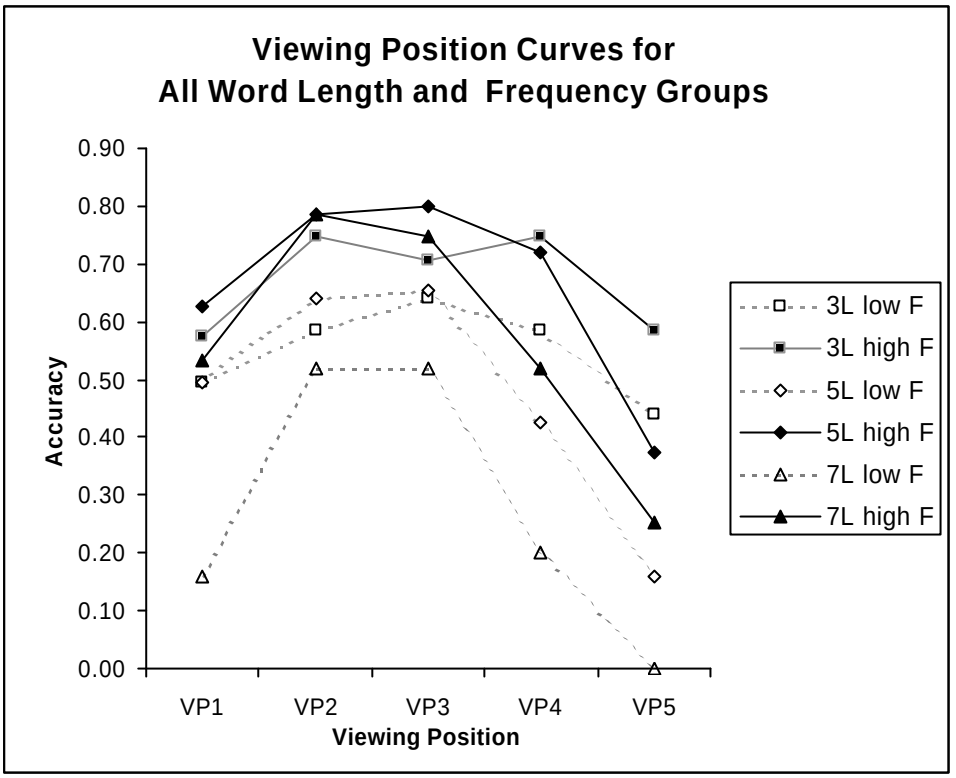
Table 6: Stimulus Characteristics of Experiment 4. Log F: Logarithm of Frequency per Million +1, N: Neighbors, N-Grad: Neighborhood Gradient that may influence the asymmetry of the curve (see Study 3 for details), but is fairly constant here.

2.2.3. Procedure

The experiment was run on an Apple Macintosh Powerbook 190CS. The procedure was identical to Experiments 1 and 2 except for the mask consisting of 9 '#' signs, regardless of word length. Thus, the mask gave no indication about the word length. Participants received 10 training trials in the training session and five buffer trials in the experimental session prior to the randomized 150 experimental trials. Participants were not informed about the word lengths used in the experiment. They were instructed to respond as accurately as possible. The whole experiment took about 20 minutes.

2.3. Results

Figures 15, 16, 17, and Table 7 show clear main effects of viewing position ($F(4, 56) = 45.29$, $p < .001$), frequency ($F(1, 14) = 126.93$, $p < .001$), and word length ($F(2, 28) = 11.46$, $p < .001$), and two interactions: Between word length and viewing position ($F(8, 112) = 5.27$, $p < .001$) and word length and frequency ($F(2, 28) = 10.88$, $p < .001$). While Figure 15 depicts the raw data for each condition, these interactions can be more closely examined in Figures 16 and 17. Optimal viewing position for five-letter words was, as always, slightly left of the center. This asymmetry disappeared and almost tended to become reversed for shorter words, while for seven-letter words it was similar to five-letter words (see also model fits). Finally, the frequency effect clearly increased with word length (Figure 17).



Standard Errors for Raw Viewing Position Curves

# Letters	Freq	VP1	VP2	VP3	VP4	VP5
3	low F	0.08	0.08	0.09	0.08	0.09
3	high F	0.09	0.06	0.08	0.07	0.09
5	low F	0.08	0.07	0.08	0.07	0.06
5	high F	0.06	0.05	0.06	0.07	0.06
7	low F	0.04	0.07	0.08	0.06	0.00
7	high F	0.07	0.05	0.05	0.08	0.06

Figure 15 and Table 7: Viewing position curves for all word length and frequency groups. Low-frequency curves are dotted lines and high-frequency curves are continuous lines. Standard errors were omitted in this graph to retain a good graphical display of the results, but are given in Table 7. A main frequency and a main word length effect and their interactions with viewing position can already be observed. Longer and less frequent words tended to suffer particularly from non-optimal viewing positions.

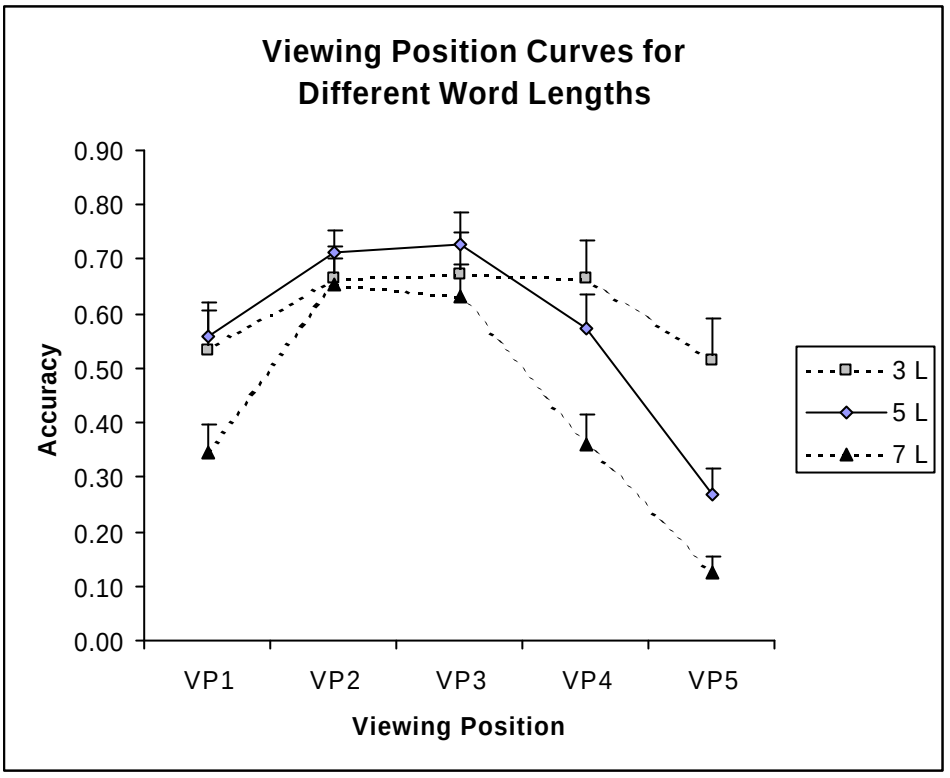


Figure 16: Viewing position curves for different word lengths with standard error bars depicting the word lengths by viewing position interaction. While near the optimal viewing position (i.e., positions 2 and 3) almost no word length effects could be observed, the word length effect increased for inferior viewing positions 1, 4, and 5.

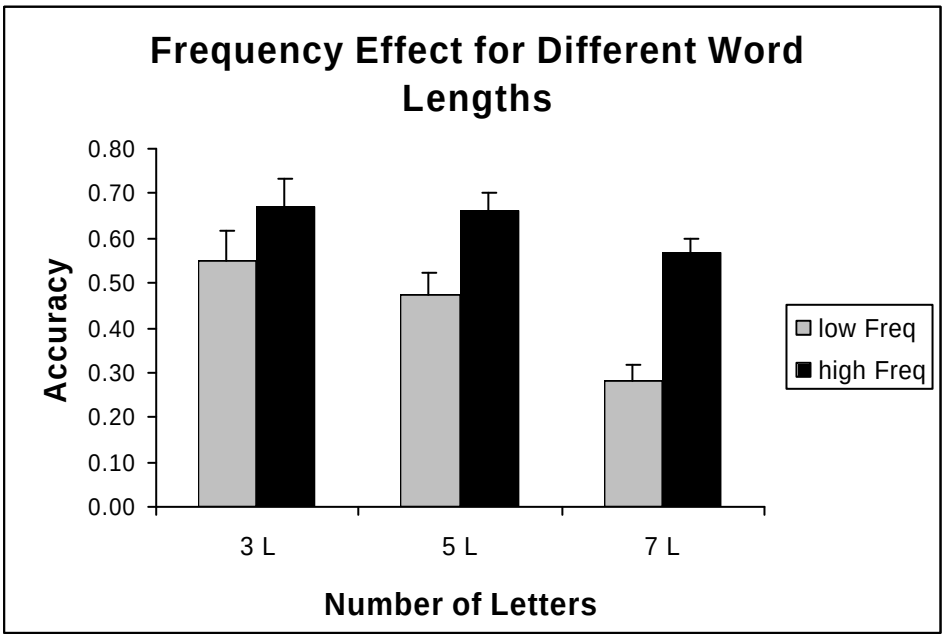


Figure 17: Frequency by word length interaction over all viewing positions with standard error bars. The frequency effect increased with word length.

More importantly, the word length effect was a function of viewing position. I compared central viewing position with mean performance on the edge viewing positions, i.e., final and beginning viewing positions for all frequency groups. In a 2×3 ANOVA for central vs. edge viewing positions over all three word lengths, I found an interaction between these two viewing positions and length ($F(2, 28) = 9.26$, $p < .001$) in addition to the two main effects ($F(2, 28) = 5.71$, $p < .01$ for word length, and $F(1, 14) = 59.49$, $p < .001$ for central vs. edge viewing position). While for edge viewing position I observed a linear decrease of accuracy of 0.07 for every additional letter (linear trend: ($t(43) = -4.01$, $p < .001$), I observed no significant linear decrease of accuracy with word length for central viewing position (decrease of accuracy of 0.01 for every additional letter; linear trend: $t(43) = -0.44$; $p = 0.33$). Even when no linear trend between word lengths was assumed, the results were straightforward: On edge viewing positions, longer words produced less accurate responses than shorter words in every single statistical comparison (all $t(14) = 2.09, 4.79, 4.88$, all $p < .05$). In contrast, on central viewing position any comparison failed to reach significance ($t(14) = -0.77, 0.58, 1.48$, all $p > .05$). The implications are clear. Had I only looked at central viewing position I would not have observed a word length effect. This can also be seen in the single participants analysis of word length effects on central and edge viewing positions (see Figure 18).

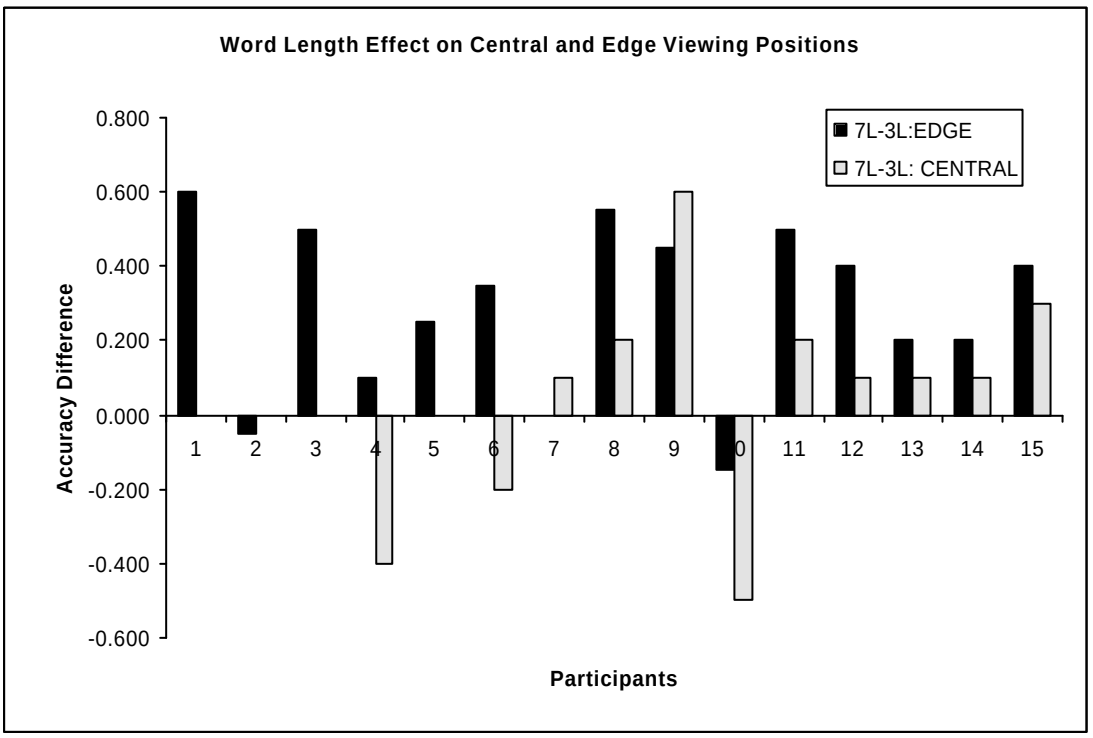


Figure 18: Differences in accuracy between seven-letter words and three-letter words on central vs. edge viewing positions. In edge viewing positions, the word length effect was prevalent in 12 out of 15 participants. In central viewing position, the overall word length effect was much weaker and could only be observed in eight out of 15 participants.

While performance for seven-letter words was inferior to performance for three-letter words on edge viewing positions in 12 out of 15 participants, on central viewing position an equally consistent word length effect could not be observed (8 out of 15 participants). Finally, if absolute rather than relative distance from fixation position was plotted, the difference between three- and five-letter words disappeared. However, seven-letter words seemed to be still less accurately recognized right from fixation than the two other word lengths (Figure 19).

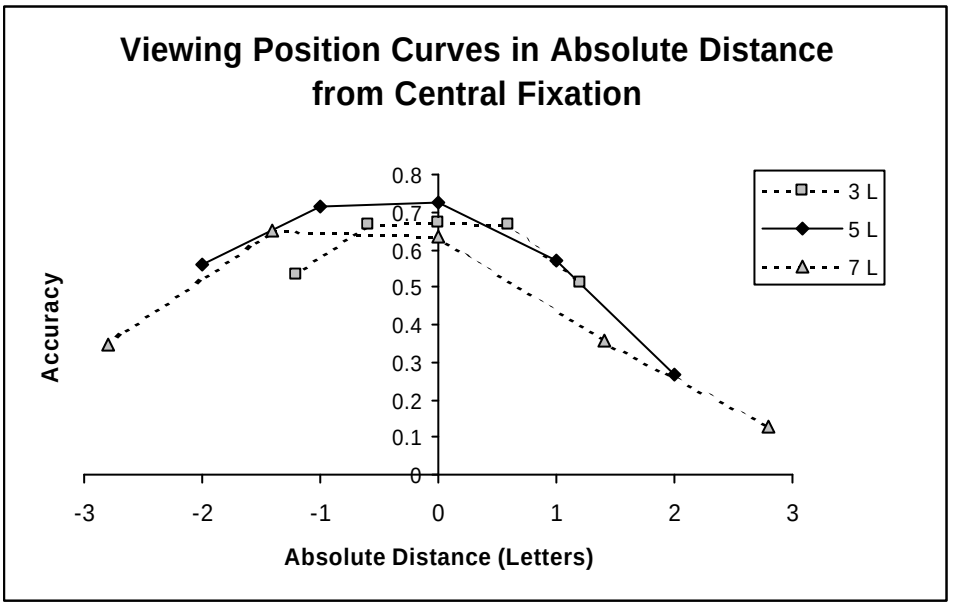


Figure 19: Viewing position curves for different word lengths in a mixed viewing position scale. From the central viewing position (a relative viewing position, which is a function of word length) absolute letter positions (i.e., independent of word length), are depicted on the abscissa. Seven-letter words were still not as well recognized, particularly when fixation was right of the center.

In exploratory data analyses, I studied in more detail what led to the frequency effect. All conditions that worsened visual word recognition performance in the current experiment led to greater frequency effects. This observation can be illustrated by directly plotting mean accuracy (i.e., averaged over frequency groups) for each viewing position and each word group against the frequency effect on that very viewing position (see Figure 20). The results are straightforward. As overall recognition decreased, the frequency effect became larger ($r = -.51$, linear trend: $t(13) = 2.14$; $p < .05$)[10]. These results also held when the accuracy values ($p \in [0\ 1]$) were transformed into z-space by inverse normal transformations ($r = -.53$, linear trend: $t(13) = 2.42$, $p < .05$; for $\Phi^{(-1)}(0) = -2$; and $r = -.59$, linear trend: $t(13) = 2.64$; $p = .01$ for

[10] Note that in a random sample for X_1, X_2 $U([0,1])$ i.i.d. the $E\ |X_1-X_2|$ is maximal for $E(X_1) = E(X_2) = 0.5$. One would expect an inverse U-shaped curve that drops off to both sides from $p = 0.5$. Clearly, that was not the case.

$\Phi^{(-1)}(1) = -3)$[11]. The above results could further be confirmed by a direct analysis of the dependent variables. If a frequency effect really would increase under inferior presentation conditions, then it should become larger with increasing distance from central fixation. This seemed to be the case, both with regard to relative viewing position (linear trend: $t(43) = 4.46$; $p < .001$) and with regard to absolute distance from fixation (linear trend: $t(103) = 7.49$; $p < .001$). Finally, because words are usually not as well recognized to the right of central fixation compared to the left of central fixation, the frequency effect should be greater to the right. Again, this could be observed ($t(14) = 5.99$, $p < .001$). In sum, all explorative analyses in the current study indicated that the frequency effect increased as a function of visual variables, such as absolute and relative distance from fixation, and side of fixation. Moreover, when the frequency effect was examined as a function of overall performance over all presentation conditions, it became larger under inferior presentation conditions.

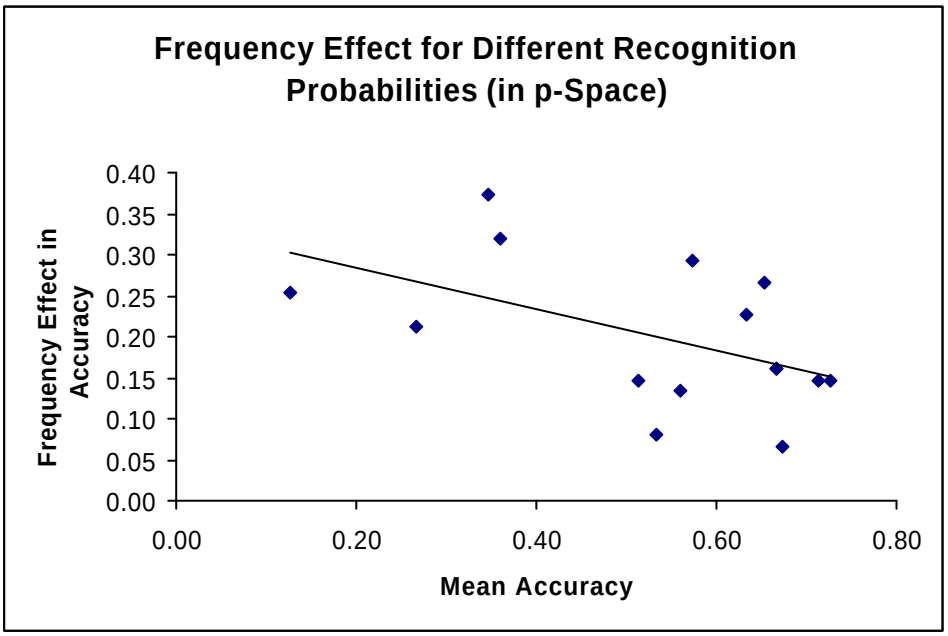


Figure 20: Direct performance analysis of the frequency effect. On each viewing position and for each word length, mean performance was computed and plotted against the frequency effect in these conditions. The frequency effect increased on viewing positions with poor performance.

[11] Φ depicts the distribution function of the standard normal distribution: $\Phi \sim N(0,1)$.

Finally, a word length by frequency interaction was obtained on edge viewing positions, but not on the central viewing position (Figure 21; $F(2, 28) = 4.59$; $p < .05$ on edge, $F(2, 28) < 1$ on central viewing position). On edge viewing positions, the clearest word length effect was observed for low-frequency words. Besides an overall word length effect, the interaction between word length and frequency was also systematic on edge viewing positions: The frequency effect was larger for longer words. In contrast, on the central viewing position neither a main word length effect nor a systematic interaction could be observed. Thus, not only these hallmark effects themselves, but also the detection of their interaction may heavily depend on the viewing position. If stimuli had only been presented on central viewing position, neither a word length effect itself nor a word length by frequency interaction would have been observed. Moreover, the obtained frequency effect would have been smaller than over all or over edge positions.

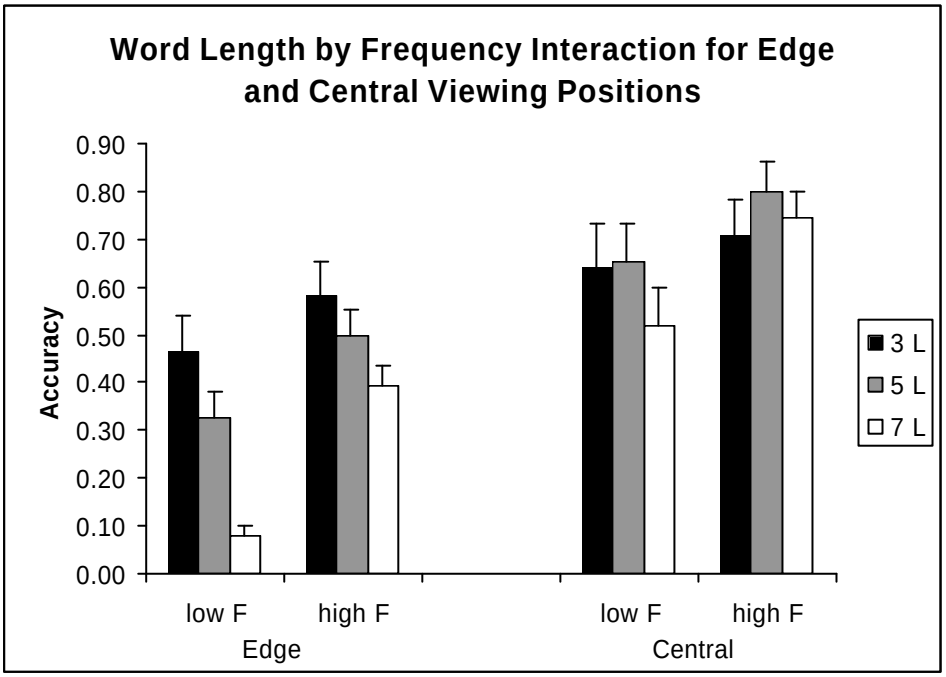


Figure 21: Word length by frequency interactions for central and edge viewing positions (i.e., positions 1 and 5). While a clear word length effect and a clear interaction with frequency could be observed for edge viewing positions, no such effect could be observed for central viewing position.

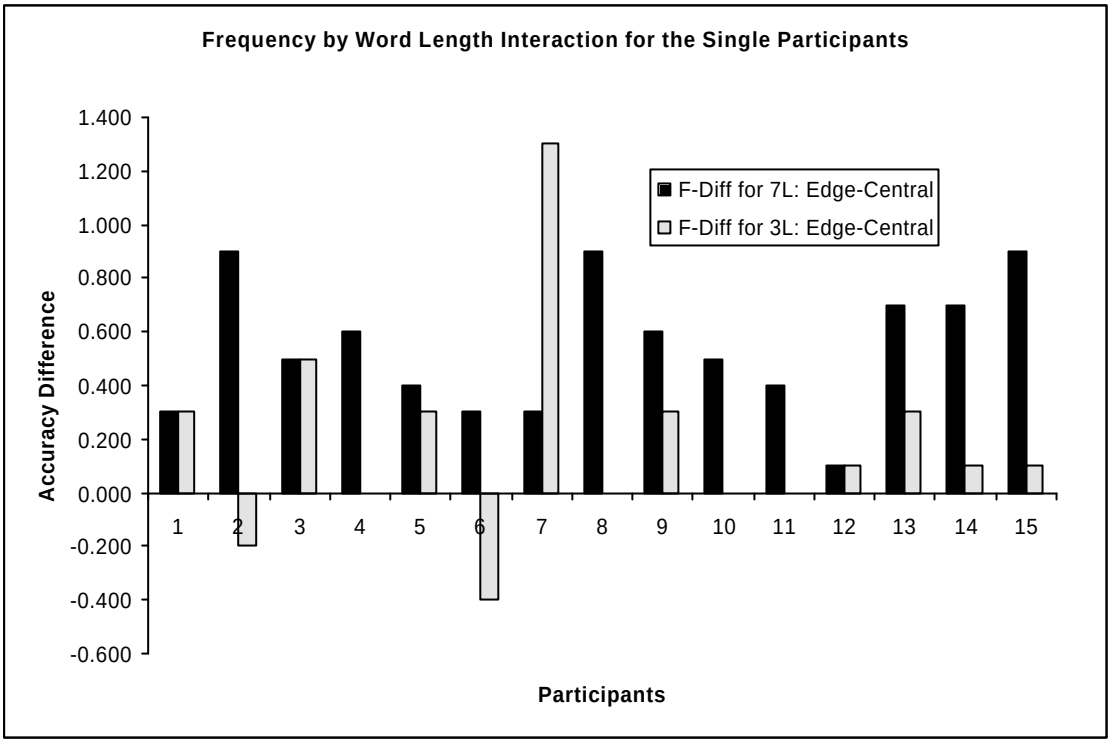


Figure 22: The difference in the frequency effect between edge and central viewing positions was computed for the three- and seven-letter words for each single participant. The overall positive values indicate that the frequency effect was generally larger for edge viewing positions. This was particularly the case for seven-letter words, but not as much for three-letter words

This pattern of results was also observed for the single participants. Figure 22 shows the difference in the frequency effect between edge and central viewing position for each participant. The positive values indicate that the frequency effect was greater on edge positions for almost every single participant. In particular, this was the case for all participants for seven-letter words. In contrast, for three-letter words only nine of 15 participants showed a larger frequency effect on edge viewing positions.

Hence, these results strengthen the hypothesis that both hallmark effects of visual word recognition, the frequency effect and the word length effect, systematically vary as a function of visual presentation condition. Both effects became larger as visual salience of stimulus presentation became poorer. Even their interaction seemed to increase under visually degraded conditions. Thus, at least in PITs the size of word length and frequency effects and their interaction obtained at central fixation position cannot always be generalized. In particular, these results question the general validity of null effects on central fixation because in this study null effects of word length and null interactions between word length and frequency were observed on central fixation position, but not on other viewing positions.

2.4. Model Fits

The fits of the generalized mathematical model of Nazir et al. (1998) complemented and confirmed the empirical findings, and demonstrated the capabilities and the constraints of the current model in three respects (see Figures 23, 24, and Table 8). First, the model fits showed that the asymmetry of the viewing position curve shifts as a function of word lengths, and thus confirmed the perceptual frequency hypothesis of Nazir et al. (1998). Second, the fitted curves for longer seven-letter words were too flat, while this was not the case for three- and five-letter words (see Table 8 for RMSDs). For a future model refinement, this implies that a lexical (frequency) parameter should be incorporated in such a way that its effect systematically varies as a function of word length and visual salience. A simple additive height parameter that does not vary as a function of these variables cannot be sufficient. Third, the visual drop-off rates did not only differ between high- and low-frequency groups (which is natural given that no frequency parameter is yet incorporated), but also between different word lengths. Thus, the assumption of perfect recognition of a given letter on fixation does no longer hold when inferior performance for longer words should also be fitted with invariant visual drop-off parameters.

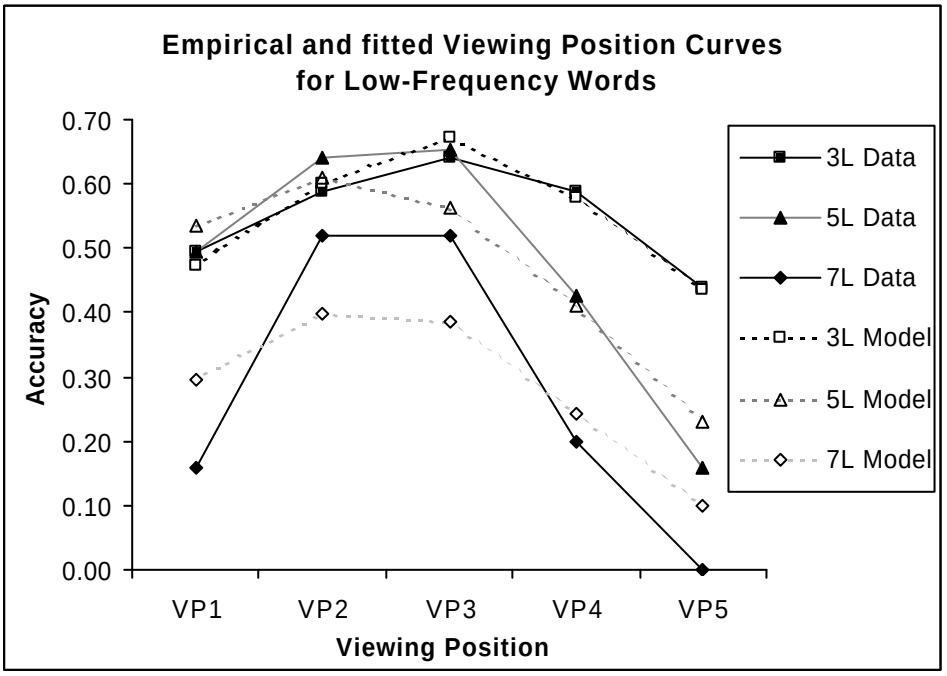


Figure 23: Empirical (continuous lines) and fitted viewing position curves (dotted lines) for different word lengths for low-frequency stimuli. The model without frequency parameter underestimated the drop-off for long words.

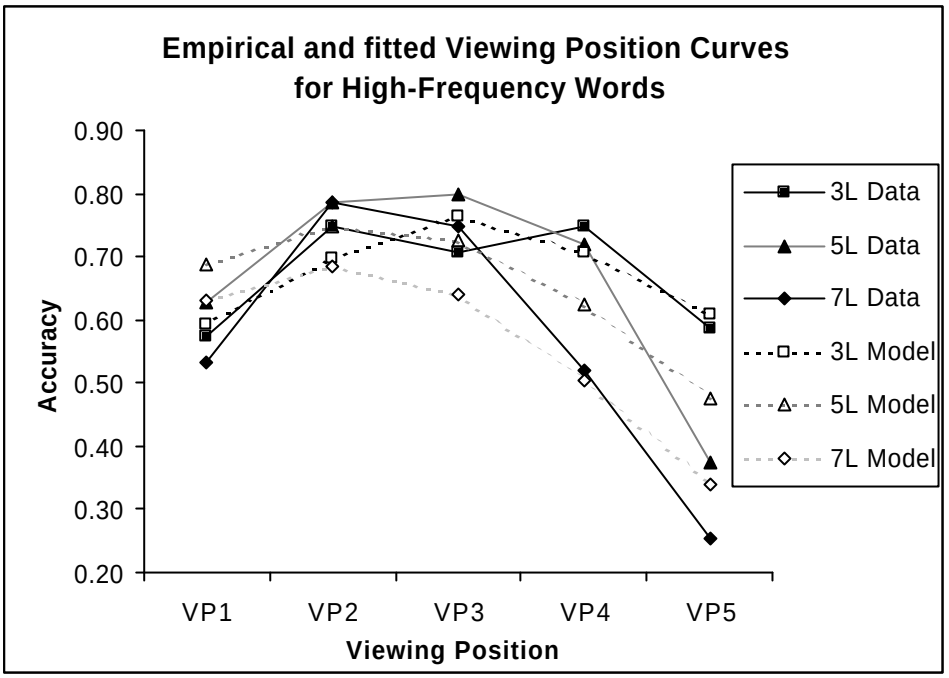


Figure 24: Empirical (continuous lines) and fitted viewing position curves (dotted lines) for different word lengths for high-frequency stimuli. The model without frequency parameter underestimated the drop-off for long words.

Model Parameters and RMSDs				
# Letters	**Freq**	d	A	RMSD
3	**all F**	0.15	1.04	0.027
5	**all F**	0.05	1.93	0.062
7	**all F**	0.04	1.86	0.092
3	**low F**	0.17	1.09	0.017
3	**high F**	0.13	0.96	0.041
5	**low F**	0.06	2.06	0.056
5	**high F**	0.04	1.87	0.079
7	**low F**	0.05	1.70	0.112
7	**high F**	0.02	2.24	0.091

Table 8: Parameters and root mean square deviations for the model fits. Freq, F: Frequency, d: Drop-off parameter to the right, A: Asymmetry, RMSD: Root Mean Square Deviation between empirical and fitted curve. The fits for the three-letter words were almost symmetrical, while the fits for five- and seven-letter words were heavily asymmetrical. Drop-off rate for three-letter words was much higher than for five- and seven-letter words, but the goodness of fit as measured by RMSD decreased with word length (see text for interpretation).

First, the asymmetry parameter for three-letter words was about 1 and, hence almost symmetrical, while for five- and seven-letter words that parameter was in the range of Nazir et al. (1998) between 1.70 and 2.24. These fitted asymmetry parameters correspond to the frequency of landing positions as a function of word lengths, and hence confirm the perceptual frequency hypothesis of Nazir et al. (1998).

Second, as can be well seen in Figures 23 and 24, the fitted curves were generally too flat for long words. The model underestimated the general drop-off in performance by viewing position. This was particularly the case for long words on bad viewing positions, i.e., in the least optimal presentation conditions. The reason is that the drop-off parameter influences the concavity of curve shape as well as the overall height of the curve. So, if a steeper drop-off parameter were used by the model, the overall height would be so low that the fit would be even worse than it is now because the curve was already close to the floor. Adding a frequency parameter that increases the height of the curve would allow the curve shape for long words to fall off more steeply towards the edges of the word.

This non-linear and non-monotonic interaction between the drop-off rate and the curve shape in the current form of the model again undermines a linear additive approach to incorporating lexical parameters. One could argue that if frequency particularly helps on bad positions the model should underestimate performance on bad positions rather than overestimate them. However, this argument would not take into account the non-linear interaction between drop-off rate and curve shape. Without a frequency parameter, the model does not by its very nature of construction allow for a difference of $p = .52$ in accuracy between optimal and least optimal viewing position. Only an adequately incorporated frequency parameter would allow such a steep curve shape. Note that for short words the model captured the data well. As empirically depicted in the word by frequency interaction, frequency had a lesser effect on short words, possibly because visual saliency was relatively good on all viewing positions. Thus, simple additivity of the effects of visual drop-off in legibility and of the influence of lexical frequency may not be sufficient to account for the possibly non-linear interaction between these two factors.

Third, the drop-off rate for the three-letter words was rather high compared to words with more letters in other experiments (Nazir et al., 1998; see also Experiments 1 and 2 and later Experiment 6.). The model produced such a high drop-off rate because the letter on fixation (or the letter that would be on fixation) is assumed to be recognized with probability 1. If one of three letters is recognized with probability 1, the model needs to assume an unusually high drop-off rate to account for an overall inferior performance because the word recognition probability is computed as the product of the single letter

recognition probabilities. The two remaining letter recognition probabilities can only be lowered strongly by an unusually high drop-off rate. Therefore, the model cannot (and did not) produce reasonable and fairly invariant drop-off rates for three-letter words, while it did do so for longer words. This consideration suggests that the simplifying assumption of fixing recognition probability at 1 does not work for shorter words and needs to be changed in a more general parameter for a functional model for any word length. Clearly, if the visual drop-off parameter should be invariant for different word lengths the current form of the model would not allow good fits.

To summarize, the fits may allow the following implications: Because the asymmetry parameter changed as a function of word length, the model fits strengthen the perceptual frequency hypothesis of word length. Second, the fitted curves for long words were too flat, implying that a frequency parameter must be added to allow steeper drop-offs of performance towards the edges. Third, as the drop-off rates for short words were exceptionally high, the simplifying assumption that letter recognition likelihood on fixation is 1 must probably be given up.

2.5. Discussion

Experiment 4 revealed four main results. First, there was a systematic interaction of frequency with viewing position. The frequency effects became larger as viewing position became more inferior. This was indicated by two different analyses. The frequency effect increased as a function of physical distance from central fixation, and, even more directly, the frequency effect increased as a function of overall performance per viewing position. The latter result was probably not due to a scaling problem, since an analysis in probability space and in z-space yielded virtually identical results. Second, word length effects increased systematically with inferior viewing position. While for central viewing position, no word length effect could be obtained, for the word edge viewing positions (i.e., first and last viewing position) a clear word length effect could be obtained. If a mixed procedure was used to investigate viewing position, i.e., absolute letter position distance from the relative central viewing position, then the word length by viewing position interaction still affected seven-letter words. However, in this most unfavorable condition for detecting differences (see general introduction), three- and five-letter words did not differ much. Third, there was a systematic interaction between word length and frequency. The interaction disappeared for central viewing position, but was huge for the edge viewing position. Fourth, model fits revealed that the fitted curves are particularly bad for longer words. This is what would be expected from a model without a lexical parameter if one assumes that frequency exerts stronger top-down effects on the recognition of

longer words via this lexical parameter (which is suggested by the observed interaction of frequency and word length). As long as the frequency parameter is not incorporated and the overall height of the curve is determined by the drop-off parameter only, the drop-off parameter is biased towards fitting overall height of the curve. In conditions in which the influence of the frequency parameter is particularly strong, this lack of a frequency parameter may particularly affect the goodness of fit. The qualitative nature of these results, however, indicates that the model of Nazir et al (1998) could be refined according to the nested modeling principle (cf. Jacobs & Grainger, 1994).

With regard to word recognition models three things seem important for this thesis. First, the result that word frequency interacted with visual legibility of a word seems to be in line with interactive models assuming top-down processes that take legibility and confusability of letters into account (e.g., Jacobs et al., 1998; McClelland & Rumelhart, 1981; Rastle & Coltheart, 1999), but not with models that assume separate successive stages for visual and lexical processing (e.g., Besner & Smith, 1992; Borowsky & Besner, 1993). Second, the interaction of the word length effect with viewing position is qualitatively well captured by the assumption that legibility of letters linearly decreases with distance from fixation (cf. Nazir et al., 1992; Olzak & Thomas, 1986). I believe that future models of visual word recognition modeling the word length effect should take this decrease in legibility into account.

Finally, together with the similar results of O'Regan and Jacobs (1992, for the LDT and the NT) my results have direct consequences for the interpretation of almost any experiment in visual word recognition in which frequency and word length are manipulated on central fixation position only. In the three most frequently used word recognition tasks (LDT, NT, and PIT, cf. Grainger & Jacobs, 1996), any null effect of frequency or word length and any null interaction between themselves or with other variables may not generalize from central viewing positions to other viewing positions. Thus, its interpretation as a general effect for a word recognition model may be premature because the effect may, instead, be an artificial product of laboratory condition with unlimited central fixation as an almost optimal presentation condition that rarely occurs in an ecological environment (see, e.g., Nazir et al., 1998). As this effect was found task-overlapping in the PIT and similarly, in the NT and LDT (O'Regan & Jacobs, 1992), it is likely to be more than an artificial postlexical guessing effect (cf. Paap & Johansen, 1994).

3. Experiment 5: Testing the Alteration of the Frequency Effect in a Word Fragmentation Task

3.1. Introduction

The central hypothesis of Experiment 4 was that frequency effects become stronger as visual salience of a stimulus becomes poorer. In Experiment 4, visual salience was manipulated by choosing an optimal or poor viewing condition. Therefore, one could argue that the observed interaction of frequency and visual quality of stimulus presentations is restricted to the manipulation of viewing condition, but does not generalize to other visual manipulations. Degradation or distortions of visual salience of stimuli have brought inconsistent results with some authors finding interactions between degradation and frequency, but not others (see introduction and Allen et al., 1995; Becker & Killion, 1977; Besner, 1989; Besner & McCann, 1987; Borowsky & Besner, 1993; Kinoshita, 1987; Mayall & Humphreys, 1996b; Norris, 1984; Wilding, 1988). Therefore, I investigated the interaction of frequency with visual salience again in a recently published fragmentation task (Snodgrass & Mintzer, 1993; Snodgrass & Poster, 1992). I chose this fragmentation task because a regression analysis demonstrated that it is most sensitive to frequency and visual confusability of the single letters as compared to other variables of visual word recognition (Ziegler et al., 1998). Thus, it may offer a good chance to detect an interaction between the two variables as they explain most of the variance in this task. According to interactive activation models and my above hypothesis, frequency should interact with visual letter confusability, and particularly influence performance for visually highly confusable, not well legible, words. In contrast, if visual quality and frequency affect sequential stages no interaction should be observed. In sum, this investigation represents a cross-validation of the observations of Experiment 4.

3.2. Method

3.2.1. Participants

10 psychology students of the Philipps-University Marburg participated in the experiment. All were native German speakers and had normal or corrected to normal vision. They received course credit for participation in the experiment. None of them had participated in Experiment 4.

3.2.2. Stimuli and Design

80 German lemmas from the CELEX-Database (Baayen, et al., 1993) were used in Experiment 5. All stimuli were 4-letter monosyllabic words (see Table 9 and Appendix E). They were divided into 4 subgroups consisting of 20 items each in a 2 * 2 ANOVA design (visual letter confusability * frequency). Additionally, N was controlled between all of these groups and higher frequency neighbors (HFN) between the respective frequency groups, so that any interaction cannot be attributed to differences in HFN either. The letter confusability index (LCI) was computed as in Ziegler et al. (1998) with the exception that normalization of performance was now computed over letters rather than over the stimulus sample (see Appendix F). The letter confusability index of a given word is now the average of the letter confusabilities of the single letter. In this way the letter confusability index of a target word is constant and independent from the word sample used in a given experiment.

	high Frequency (logF > 1.5; F > 31)		**low Frequency (logF< 1; F < 10)**	
	high LCI	**low LCI**	**high LCI**	**low LCI**
logF/1Mio	2.12	2.15	0.40	0.41
LCI	0.76	0.30	0.75	0.26
N	4.70	4.70	4.60	4.60
HFN	1.20	1.00	3.35	3.25
N-conf	0.09	0.06	0.08	0.05

Table 9: Stimulus Characteristics of Experiment 5. LCI: Letter Confusability Index (cf. Ziegler et al., 1998), N: Number of Neighbors, HFN: Number of higher frequency neighbors, log / 1 Mio: Logarithm of frequency per million + 1 to the base 10, N-conf: Neighborhood Confusability measure that will be used and defined in Study 3, Experiment 7 (see below).

3.2.3. Procedure

The screen fragmentation procedure was identical to that used by Ziegler et al. (1998, see Figure 25). At the beginning of the experiment the participants were familiarized with the typography by presenting all the letters of the alphabet. During the experimental trials a test word was first displayed at the most fragmented level of presentation (level 1). Participants were asked to gradually demask the word by pressing the space bar of the keyboard until they had sufficient information to generate a response. This

demasking process was entirely controlled by the participants. They were asked to give a response by typing the word on the keyboard as soon as they thought they knew what the stimulus was. They were instructed to respond as quickly and accurately as possible. Identification thresholds and errors were recorded. Responses were not corrected and, in contrast to some experiments of Snodgrass and Mintzer (1993), participants could only give one response on each experimental trial. After the response was given the next trial was automatically initiated. The fragmentation procedure is illustrated in Figure 25 and is described in detail in Ziegler et al. (1998).

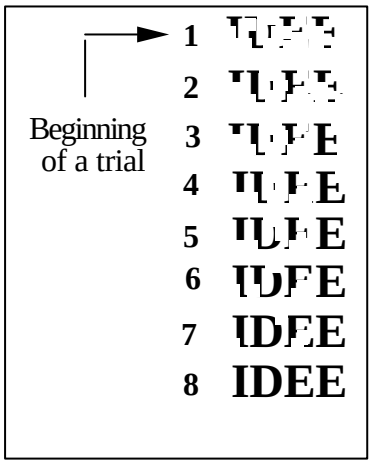


Figure 25: Levels of the word fragmentation task.

An experimental session was subdivided into 2 blocks of 40 words in random order for each participant. Prior to the experiment participants received 6 trials to familiarize them with the experiment.

3.3. Results

The results of the ANOVAs for participants (F_1) and items (F_2) for answer levels were straightforward (see Figure 26). They revealed main effects of LCI ($F_1(1, 9) = 156.22$, $p_1 < .001$; $F_2(1, 76) = 67.24$, $p_2 < .001$), frequency ($F_1(1, 9) = 5.86$, $p_1 < .05$; $F_2(1, 76) = 5.25$, $p_2 < .05$) and an interaction between frequency and LCI ($F_1(1, 9) = 5.74$, $p_1 < .05$; $F_2(1, 76) = 4.08$, $p_2 < .05$). As expected this interaction revealed that frequency had a significant effect when visual legibility was poor ($t_1(1, 9) = 3.59$, $p_1 < .01$; $t_2(1, 38) = 3.51$, $p_2 < .001$), but no significant effect when visual legibility was good ($t_1(1, 9) = 0.06$, $p_1 = .48$; $t_2(1, 38) = -.16$, $p_2 = .43$). There was no indication of a speed accuracy trade-off (SATO). There was no significant accuracy effect of frequency ($F_1(1, 9) = 1.09$, $p_1 = .32$; $F_2(1, 76) < 1$), no interaction ($F_1(1,9) < 1$; $F_2(1,76) < 1$), and only a slight trend for less confusable words to be more accurate in the participant analysis ($F_1(1, 9) = 4.57$, $p_1 = .06$; $F_2(1, 76) = 2.04$, $p_2 = .16$). Descriptively, all the accuracy effects pointed in the same direction as the recognition level data (see Figure 27). Conditions that produced faster responses tended, at least descriptively, also to produce less accurate responses. Thus, a SATO cannot account for any effect of the present data pattern. Finally, for the combination measure T% the ANOVA also revealed significant results (see Figure 28). It revealed main effects of LCI ($F_1(1, 9) = 65.99$, $p_1 < .001$; $F_2(1, 76) = 41.99$, $p_2 < .001$), frequency ($F_1(1, 9) = 26.74$, $p_1 < .001$; $F_2(1, 76) = 3.23$, $p_2 < .08$), but due to the null effect for errors the interaction between frequency and LCI did not reach significance ($F_1(1, 9) = 1.03$, $p_1 = .34$;

$F_2(1, 76) = 1.23$, $p_2 = .27$). Individually, frequency tended to affect the recognition speed for highly confusable words more than for lowly confusable words in most participants (7 of 10, see Figure 29). For highly confusable words, the frequency effect also seemed to be more consistent across participants (9 out of 10) than for lowly confusable words (5 out of 10, see Figure 29).

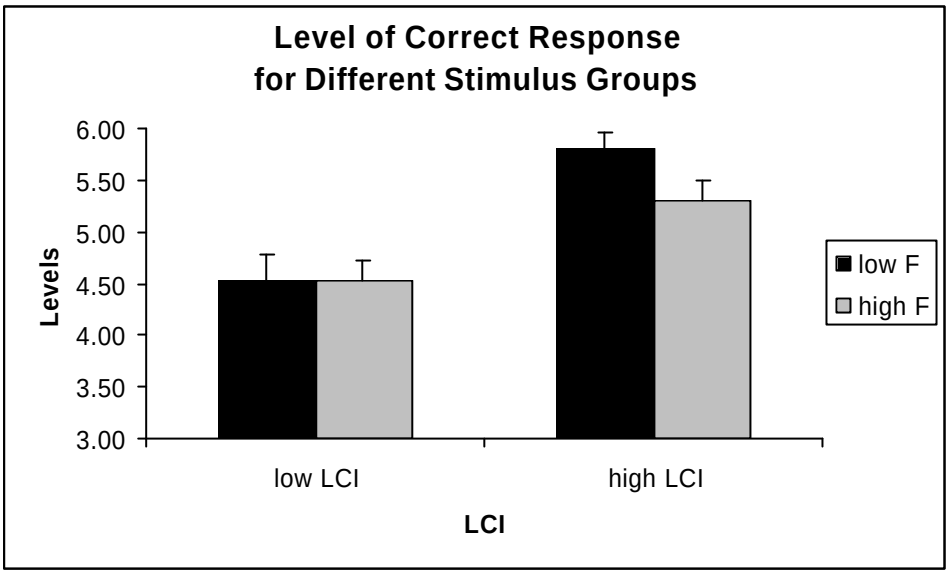


Figure 26: Mean level of correct response with standard error bars for different stimulus groups in Experiment 5. The frequency effect was larger under the non-optimal presentation condition, i.e., with high letter confusability.

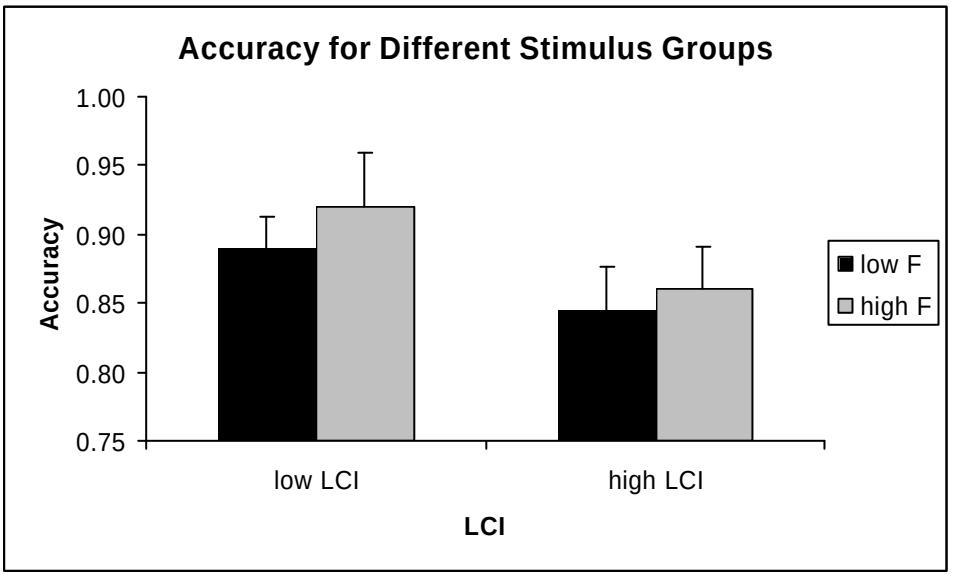


Figure 27 : Means and standard errors for accuracy for different stimulus groups in Experiment 5. No speed-accuracy-trade-off could be observed.

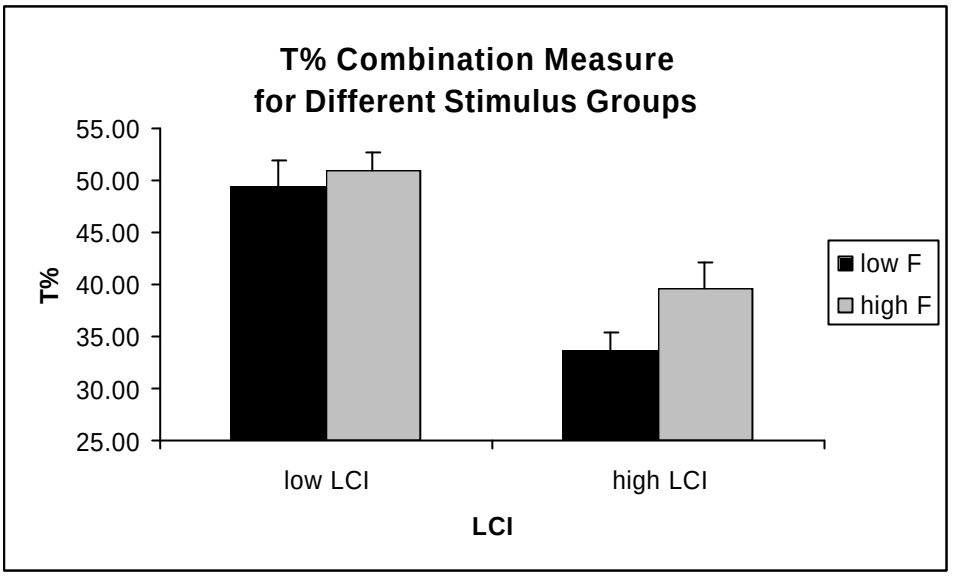


Figure 28: Means and standard errors for the T% combination measure for different stimulus groups in Experiment 5.

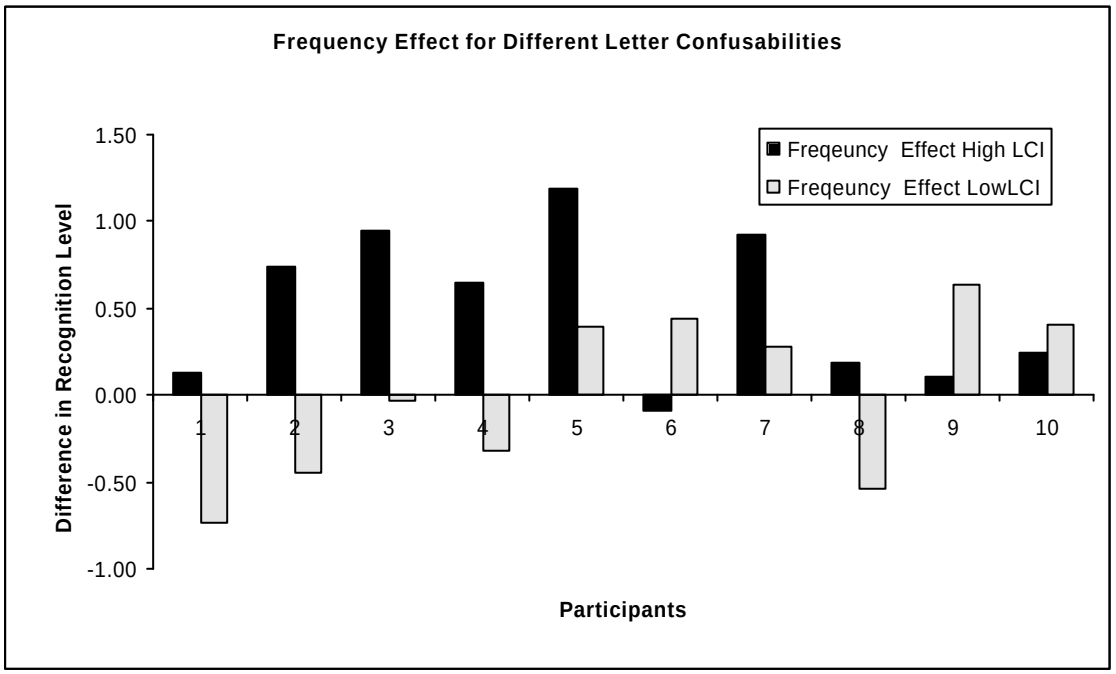


Figure 29: Frequency Effects for high and low letter confusabilities for each single participant. Frequency effects for high LCI could descriptively be observed in 9 out of 10 participants, but only in 5 out of 10 participants for low LCI.

3.4. Discussion

In Experiment 5, both frequency and visual letter confusability affected performance in the word fragmentation task. Most importantly, they interacted in the expected direction. High frequency improved performance more for badly legible words. None of these results can be attributed to a SATO between

conditions as accuracy data were either in the same direction (i.e., faster responses were more accurate) or non-significant. The results from Experiment 4 can thus be generalized in several respects. They are neither specific to the type of perceptual manipulation used in Experiment 4, nor to the choice of the dependent variable (accuracy) because in Experiment 5 I obtained an interaction effect in recognition level. Finally, Experiment 5 replicated Experiment 4 with completely different stimuli (4 letter words instead of, three-, five-, and seven-letter words). Therefore, the interaction between frequency and visual quality is also not due to specific stimulus characteristics.

4. Discussion of Study 2

Frequency effects systematically interacted with visual salience of the presented stimuli in my study: In two PITs frequency affected performance most when visual legibility was poor and bottom-up information was not salient. This interaction corroborates an interactive activation account of visual word recognition (Jacobs et al., 1998; Plaut et al., 1996; Rastle & Coltheart, 1999), but seems to be inconsistent with models that assume strictly sequential stages of visual and lexical processing (Besner & Smith, 1992; Borowsky & Besner, 1993). The effect was found task-overlapping in two different PITs and was neither due to the dependent variable (accuracy vs. level of recognition), nor to visual manipulation (letter confusability vs. viewing position) nor the specific stimuli attributes for selection. Similarly, word length interacted with viewing position. Word length effects were larger when viewing position was poorer. Moreover, the interaction of frequency with word length was greater when visual legibility was poor.

These results have three main implications that will be addressed in the general discussion in greater detail: First, null effects of or null interactions with frequency obtained under near-optimal laboratory presentation conditions are not conclusive. Using additive stages logic to argue that frequency and other visual processes are occurring on subsequent independent stages may be a pitfall if presentation is restricted to near-optimal conditions.

Second, the results impose important constraints for a refined mathematical viewing model at the end of this: A future model needs a lexical parameter, whose influence must not be just additive for all visual conditions, but stronger for longer words and for poorer viewing positions. Furthermore, for longer words the fitted curves were too flat compared to the empirical curves. However, if the drop-off for longer words were much steeper than it is now, one would obtain negative recognition values on edge positions. To avoid a model fit using rather low drop-off rates (and consequently producing rather flat curves), a refined model should be constructed in such a way that negative recognition probabilities are no longer possible.

Finally, if one wants invariant visual drop-off rates for different word lengths, the simplifying assumption that recognition probability at fixation is 1 may not be appropriate anymore (see Experiment 4 for details).

Third, the results are consistent with both perceptual learning hypotheses of location and word form: The shift in asymmetry with word length follows the perceptual frequency of the fixation (landing) positions in natural reading for different word lengths. Together with Study 1, these results underline the appropriateness of the perceptual frequency hypothesis for visual word recognition. Altogether, the study demonstrates that the influence of perceptual visual processes should not be underestimated in visual word recognition research by using the principle of isolated variation for good visual presentation conditions and interpreting resulting null interactions using AFL.

IV. Study 3: Neighborhood as an Index in Visual-Orthographic-Lexical Space

1. Introduction

Investigating the influence of visual determinants on orthographic-lexical processes presents a new step in this thesis. While in the previous chapters individual properties of a word in a given presentation form were studied (perceptual frequency, lexical frequency, letter confusability or legibility, or word length), now the relation of a word to other words of the lexicon will be examined. Neighborhood measures (e.g. N or HFN) are usually conceptualized as similarity indices in orthographic-lexical space. The purpose of Study 3 is to investigate whether orthographic relations between words are also modulated by visual determinants. If this was the case, the notion of an orthographic-lexical space would not be sufficient to describe neighborhood effects. Instead, neighborhood measures would index similarity in visual-orthographic-lexical space.

In the general introduction I have presented evidence that certain visual determinants, such as viewing position or letter confusability, exert their effects whilst interacting with orthographic-lexical information distribution within words. In particular, I have discussed one study by Grainger et al. (1992), who manipulated fixation position, neighborhood position (either letter position 2 and 4), and the number of higher frequency neighbors in two LDTs. Most important for the present study was that Grainger et al. (1992) found no general neighborhood position effect but an interaction of neighborhood position with viewing position. However, their experiment had certain limitations (see general introduction for details). First and most important, they only investigated two viewing positions and basically found an interaction of viewing position and higher frequent neighbor position for viewing position 2 but not for position 4 in five-letter words. Possible implications for other viewing positions, particularly for the central viewing position mostly used in word recognition experiments, can only be assumed. Second, while the position of high-frequency neighbors was controlled (position 2 or 4), the position of the other neighbors appears not to have been controlled (see introduction for unexpected differences between their Experiments 1 and 2). Third, the visual manipulation was restricted to having higher frequency neighbors on and off fixation. It was not examined whether the distance to fixation plays a role. However, this factor may also play a role because the legibility of the critical letters systematically drops with increasing distance to fixation (Nazir

et al, 1992). Fourth, the findings were restricted to the LDT. Following the principle of functional overlap, it is worthwhile to know whether a certain effect is task-specific or task-overlapping.

Study 3 was constructed to tackle these issues. I continuously manipulated viewing position and did not restrict the stimulus selection to words with neighbors only on certain positions (e.g., positions 2 or 4). No such restricted criterion was set. Instead, I selected words that had most of their neighbors (and thus their information ambiguity) at the beginning of the word and compared them with words that had most of their neighbors at the end of the words. Moreover, the position of any neighbor was manipulated rather than the position of higher-frequency neighbors only.

As regards the model of Nazir et al. (1998), performance should be relatively better on those fixation positions where information ambiguity (i.e., neighborhood positions) is mostly located, if there is an interaction between orthographic information ambiguity and letter legibility. The curves for words with most neighbors at the beginning and those with their neighbors at the end should differ. In particular, the asymmetry ratio should differ such, that words with their information ambiguity at the beginning produce more asymmetrical curves than words with their information ambiguity at the end (see general introduction for details). However, one needs a computational index that numerically specifies what is exactly parametrically meant by “most neighbors at the beginning” and can be manipulated in a categorical design. Such an index was developed for Experiment 6 and will be specified now.

2. Experiment 6: Visual Effects of Neighborhood Distribution

2.1. Introduction

Experiment 6 was constructed to investigate whether a systematic interaction between the neighborhood distribution and visual determinants, in particular between neighborhood position and viewing position, exists. To generalize the results of Grainger et al. (1992) to all viewing positions and all neighborhood distributions, an index is needed that characterizes any possible neighborhood distribution, but is simple enough to be manipulated in a design for which a statistical model exists. In particular, the index should provide information on how many neighbors are at the beginning or at the end of a word. The neighborhood distribution index I used was developed for this experiment.

2.1.1. The Neighborhood Gradient Index N-Grad

To compute the neighborhood index N-Grad the following procedure was used. As a base I used the empirical positional neighborhood (size) function. On the abscissa are the letter positions of a stimulus

word (see Figure 30), on the ordinate are the number of neighbors on each position. To maintain a simple index, I just computed the straight line over the positional neighborhood distribution that has the smallest root mean square deviation. This straight line is based on two parameters: The gradient which determines how steep the straight line rises or falls and the distance from the abscissa (x-axis, neighborhood positions) measured at the ordinate (y-axis, number of positional neighbors) which determines where the straight line intersects with the ordinate. The parameter determining the gradient is positive when many neighbors are at the end of the word, because then the straight line that fits the positional neighborhood distribution best should be high at the final positions and low at the initial positions. Vice versa, the gradient is negative if most neighbors are at the initial positions (see Figure 30). This gradient is called N-Grad and will be manipulated in Experiment 6.

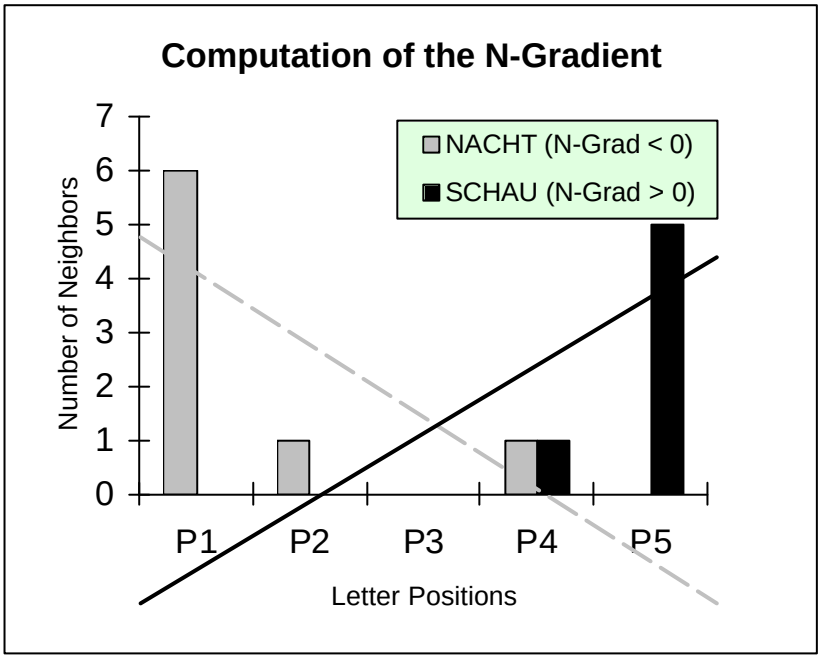


Figure 30: The computation of the neighborhood gradient N-Grad: N-Grad is the gradient of the straight line that describes the neighborhood distribution best (see text for details). In this example, N-Grad for "NACHT" is negative as "NACHT" has most of its neighbors at the initial position, while N-Grad for "SCHAU" is positive as "SCHAU" has most of its neighbors at the final position.

The reduction of the neighborhood distribution to one parameter N-Grad is necessary, because in any factorial experimental design, one independent variable is needed that can be systematically manipulated for different item groups. However, it is important to note that a reduction of any given empirical distribution - with as many data points as word length - to one single parameter is always trivially associated with some loss of information even if other important information is preserved. The loss of

information here, in particular, refers to an N-Grad of 0. N-Grad = 0 does not indicate whether there is an equal, possibly high, number of neighbors at the beginning and the end or whether there are only neighbors in the middle position. However, as the N-Grad manipulation in this Experiment is positive vs. negative this particular loss of information seems to be acceptable and adequate with respect to the purpose of this experiment.

2.1.2. Objectives

The purpose of Study 3 is to examine whether N is better understood as a visual-orthographic-lexical measure rather than an orthographic-lexical measure. To examine this issue in Experiment 6, the first specific objective is to analyze the interaction between the positional distribution of orthographic neighborhood and viewing position. If neighborhood effects are systematically influenced by the interaction between letter legibility and viewing position, then word recognition should be worse when the legibility of the ambiguous letters is low. Because letter legibility drops with increasing distance from fixation (Nazir et al., 1992), word recognition should become relatively worse if the ambiguous letter is far from fixation. Thus, words with a negative N-Grad should be relatively better recognized with fixation at the beginning of the word compared to words with a positive N-Grad. In contrast, with fixation at the end of the word the opposite should be true.

With regard to the prototypical form of the viewing position the asymmetry should be modified in a characteristic way. Recall, that recognition is normally best slightly left from the middle. For words with a positive N-Grad that have their positions of ambiguity at the end, the curve should become more symmetrical because recognition should become better on the final positions where most of the neighbors are located. In contrast, for words with a negative N-Grad, the curve should become even more asymmetrical than normal, because recognition at the initial positions, where most of the neighbors of those words are, should be enhanced. The ANOVA cannot test such systematic variations for the different treatments within one factor, as it only tests whether the treatments differ in some unspecific way from one another (in their main effects or their interaction with other factors). Because it is insensitive to the specific nature of any interaction, the ANOVA is probably too conservative for testing such a specific interaction hypothesis about the relation between neighborhood distribution and viewing position. Again, a null effect in an ANOVA does not necessarily imply that there is no specific interaction between viewing position and neighborhood distribution, but only that an unspecified test on any possible interaction between these factors in an ANOVA within the general linear model does not approach significance. As

no sufficient test statistic for inference-statistical evaluation of the hypothesis is available, the fits with the model of Nazir et al. (1998) will again be helpful for its examination. The critical parameter for this examination is the asymmetry ratio A. If A is a purely visual parameter, N-Grad as an index of the distribution of information ambiguity should not influence the asymmetry of the viewing position curve. However, if information distribution within a word does influence the viewing position curve, its asymmetry should be shifted in a characteristic way. Considering that recognition on the initial positions is normally superior and the simplex algorithm fits A between 1 (for symmetrical curves) and 2 (e.g., Nazir et al., 1998, see studies 1 and 2), a negative N-Grad should lead to greater asymmetry ratios while a positive N-Grad should lead to less asymmetrical curves than normal.

Following the principle of functional overlap (Grainger & Jacobs, 1996), I employed a different task from the LDT used by Grainger et al. (1992) to investigate whether interactions of neighborhood and visual manipulations are task-specific or not. I chose the PIT for three reasons. [Neighborhood size effects] *"in PITs have been relatively consistent. Responses to degraded stimulus presentations are less accurate for low-frequency words that are similar to many words or that have a high-frequency neighbor unless the subject is allowed successive guesses at the target"* (Andrews, 1997, p. 446). Thus, the hypothesized direction of the effects is clearer than in other tasks (e.g., the LDT, cf. Andrews, 1997; Grainger & Jacobs, 1996). Second, a mathematical model for the VPE is available for the PIT (Nazir et al., 1998) while such a model does not yet exist for the NT. Third, as Snodgrass and Mintzer (1993) pointed out, latency measures are always indirect measures of lexical competition. They argued that it would be instructive for investigating lexical neighbor competition to see whether other lexical candidates are activated. If erroneous responses indeed consist of neighbors, this is an indication that these neighbors have been activated in the experiment. Many erroneous responses are produced in a tachistoscopic PIT with accuracy as a dependent variable. These errors also provide a second test of my hypothesis. If neighborhood is better understood as an index in visual-orthographic-lexical space, then the visual attributes of the neighbors of a word should also influence the type of errors. In particular, a neighbor should be more likely to be given as an erroneous response when its ambiguity position is far from fixation rather than when it is close to fixation.

To investigate these objectives I again varied viewing position within a word according to a Latin Square design (cf. O'Regan & Jacobs, 1992). As each word is shown on five different viewing positions, non-visual properties of a word should exert their influence on all viewing positions alike. In contrast, an interaction of positional neighborhood distribution with viewing position would indicate that neighborhood

size would better be understood as an visual-orthographic-lexical effect in this task. Additionally, I again employed the model of Nazir et al. (1998) to investigate whether the asymmetry parameter varies between conditions as a function of the neighborhood distribution or not.

2.2. Method

2.2.1. Participants

15 psychology students of the Philipps-University Marburg participated in the experiment. All were native German speakers and had normal or corrected to normal vision. They received course credit for participation.

2.2.2. Stimuli and Design

150 German five-letter lemmas from the CELEX-Database (Baayen et al., 1993) were used in Experiment 6 (see Appendix G). In a 2 * 3 ANOVA design (positive vs. negative neighborhood gradient and low-, middle-, and high-frequency), each group consisted of 25 stimuli. Frequency and neighborhood gradient were matched between the respective item groups. Because most neighbors of German five letter words are on the initial positions, it was not possible to match the total number of neighbors between the neighborhood gradient groups. However, this number was roughly held constant between the respective frequency groups (see Table 10). Thus, an overall worse performance of negative N-Grad would have to be interpreted with care, as it could be due to an overall inhibitory neighborhood effect in a PIT. The viewing position procedure using a Latin Square Design was identical to Experiments 1, 2 and 4.

		log Freq	N-Grad	N
N_Grad > 0	**Low Freq**	0.38	0.48	3.00
N_Grad > 0	**Middle Freq**	1.26	0.42	3.52
N_Grad > 0	**High Freq**	2.24	0.30	2.68
N-Grad < 0	**Low Freq**	0.42	-1.19	7.72
N-Grad < 0	**Middle Freq**	1.23	-0.98	6.92
N-Grad < 0	**High Freq**	2.30	-1.08	7.84

Table 10: Stimulus characteristics in Experiment 6. log Freq: Mean logarithmic (frequency + 1) to the base 10 per million, N-Grad: Neighborhood gradient, N : Number of neighbors. Frequency (F) Groups: Low : F < 10, Middle: 10 < F < 50, High: F > 50.

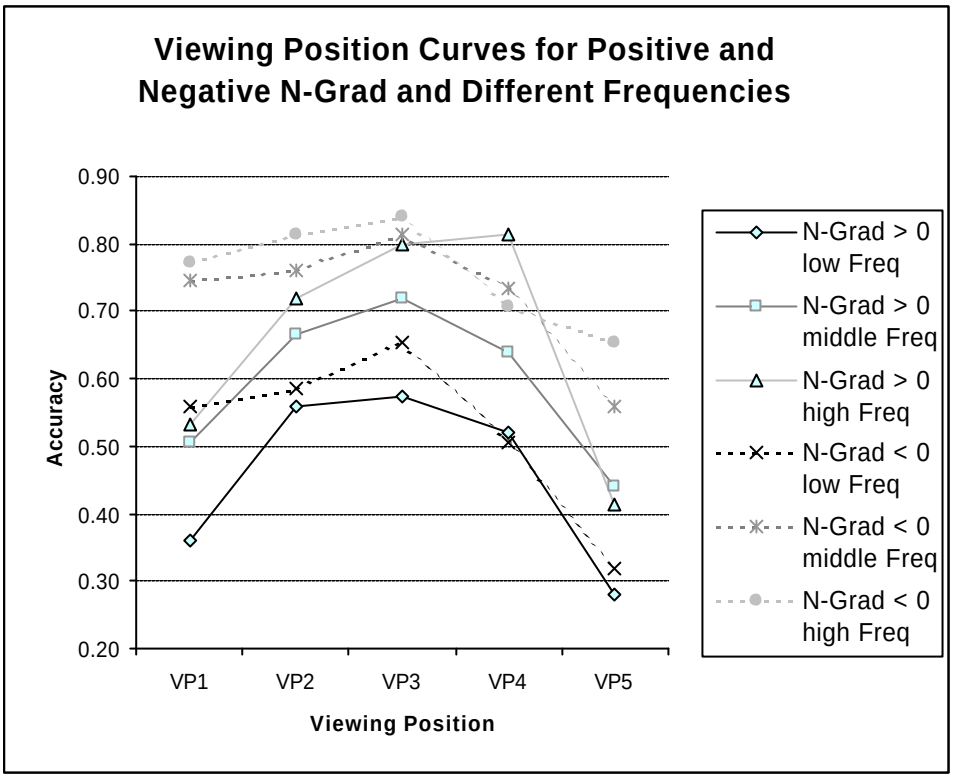
2.2.3. Procedure

The experiment was run on an Apple Macintosh Power PC. The procedure was equal to Experiments 1, 2, and 4, except for the mask consisting again of 7 '#' signs.

2.3. Results

Figures 31 and 32 and Table 11 demonstrate clear main effects of viewing position ($F(4, 56) = 21.19$, $p < .001$), frequency ($F(2, 28) = 61.34$, $p < .001$), and N-Grad ($F(1, 14) = 17.23$, $p < .001$), and an interaction between N-Grad and viewing position ($F(4, 56) = 4.90$, $p < .01$). Optimal viewing position was as expected slightly left of the center, but did shift substantially as a function of N-Grad. If N-Grad was negative and almost all neighbors were at the beginning of the target word, then performance almost did not drop at all to the left of the center. Consider the raw data of Figure 31 and the averaged curves in Figure 32: For all frequency groups, the VPE curves with negative N-Grad started almost flat, while all VPE curves with positive N-Grad dropped off to the left as viewing position curves usually do. Except for viewing position 5, which will be analyzed below, the difference between words with negative and positive N-Grad continuously decreased with increasing (i.e., more rightward) viewing position (see Figure 32).

Thus, these results suggest that if orthographic-lexical ambiguity is located at the beginning of the word, then fixating at the beginning of the words really helps recognition in a PIT. Statistically, this hypothesis has been confirmed. Words with negative N-Grad were better recognized when they were fixated at the beginning (left of the center, i.e., on positions 1 or 2) than if they were fixated at the end (right of center, i.e., position 4 and 5, $t(14) = 3.40$; $p < .01$). In contrast, for words with positive N-Grad, there was no such difference ($t(14) = 1.27$; $p = .22$, two-sided). This interaction between neighborhood distribution and viewing position could also be observed when the individual data were analyzed, although the individual data were not as consistent as in Study 2 (see Figure 33). Finally, a remarkable result was that words with a negative N-Grad produced better overall performance (see Figure 32), although in these stimulus groups N was much higher than in the positive N-Grad groups. Because N is known to be inhibitory in PITs (at least if sophisticated guessing is not allowed, cf. Andrews, 1997), this result seems to demonstrate that not the overall number of neighbors, but rather at which position they are, determines performance in the current study.



Standard Errors for Raw Viewing Position Curves

N-Grad	Freq	VP1	VP2	VP3	VP4	VP5
N-Grad>0	low Freq	0.07	0.06	0.07	0.08	0.05
N-Grad>0	middle Freq	0.08	0.07	0.06	0.06	0.07
N-Grad>0	high Freq	0.06	0.04	0.04	0.05	0.07
N-Grad<0	low Freq	0.06	0.03	0.07	0.06	0.06
N-Grad<0	middle Freq	0.06	0.05	0.05	0.06	0.04
N-Grad<0	high Freq	0.06	0.05	0.05	0.06	0.07

Figure 31 and Table 11: Raw viewing position curves for different N-Grad and frequency groups. Consider the dotted (negative N-Grad) vs. the continuous lines (positive N-Grad). Although one point was computed from only five data points per participant, for all three negative N-Grad curves (dotted lines) the usual drop-off from the middle to initial fixation position was lacking. The curves started rather flat. In contrast, for all three positive N-Grad curves (continuous lines), this drop-off from the middle to the initial fixation could be observed. In fact, all three positive N-Grad curves did not show the usual asymmetry of a viewing position curve, but were rather symmetrical (see parameters in Table 12 for a mathematical confirmation). For the sake of visibility standard errors are not included in the graph, but given in Table 11.

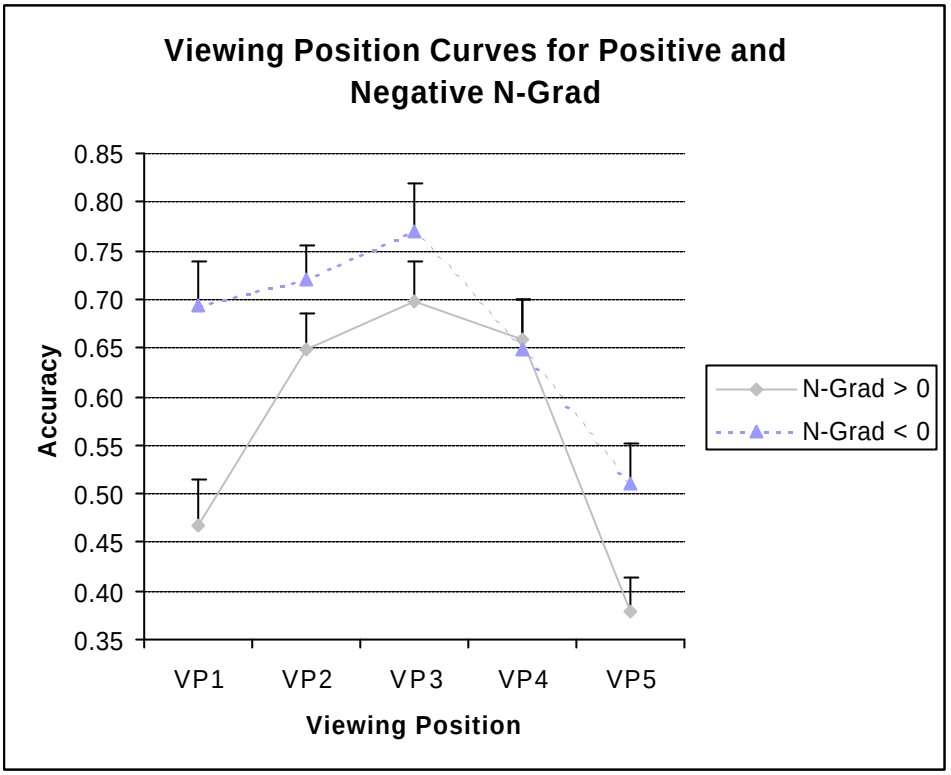


Figure 32: Viewing position curves for the two different N-Grad groups averaged over frequency. The beneficial effect for negative N-Grad was greatest on viewing position 1 and then decreased monotonically from viewing positions 1 to 4. On viewing position 5, however, words with negative N-Grad produced unexpectedly better performance again (see discussion).

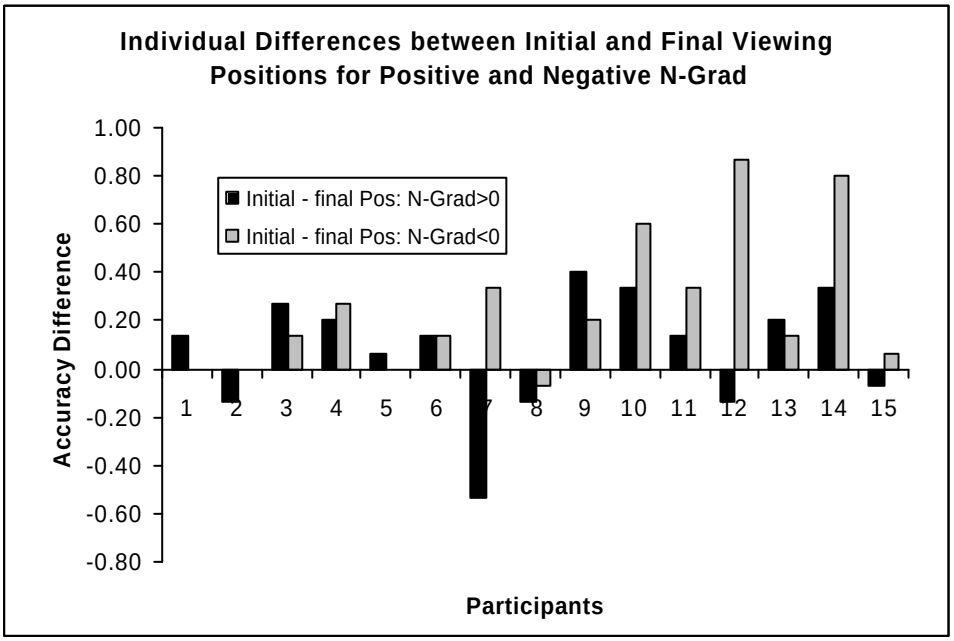


Figure 33: Words with negative N-Grad were better or at least equally well recognized on initial two viewing positions compared to the final two viewing positions in all but one participants. For words with positive N-Grad, the picture was less pronounced, but not reversed. Still, ten participants descriptively produced a little better performance on the initial two viewing positions.

The hypothesis that this shift in asymmetry is due to the neighborhood distribution can be investigated with further analyses of the erroneous responses in which lexical neighbors of the target word were given as a response. If the neighborhood gradient is an empirically meaningful concept, then words with negative N-Grad should be confused with orthographic neighbors whose letter exchange position is at the beginning of the word. To explore this issue, I computed the average letter position at which each item was confused with neighbors (words never confused with any neighbor were eliminated from this analysis). As expected, this average neighborhood confusion position was significantly and markedly more to the left in the words with negative N-Grad than in words with positive N-Grad (see Figure 34; $t = 8.27$, $p < .001$). This analysis, however, only shows that the neighborhood gradient manipulation seemed to work in general, but it does not confirm that there is an interaction between viewing position and neighborhood confusion position. To investigate this interaction directly, I coded all neighborhood confusion positions over all items and participants and regressed them on viewing position. Although each word was shown equally often on each fixation position, the analysis revealed a significant negative regression coefficient. If a word was fixated at the beginning, participants were more likely to confuse it with an orthographic neighbor towards the end of the word than if it was fixated at the end and vice versa ($r = -.24$ $p < .01$, see Figure 35). As in the overall accuracy data, position 5 again revealed unexpected results. While neighborhood error position was decreasing monotonously for position 1 to 4 for increasing viewing position, it did increase again for position 5. As in both not naturally confounded analyses viewing position 5 revealed a special unexpected effect, I explored this effect further.

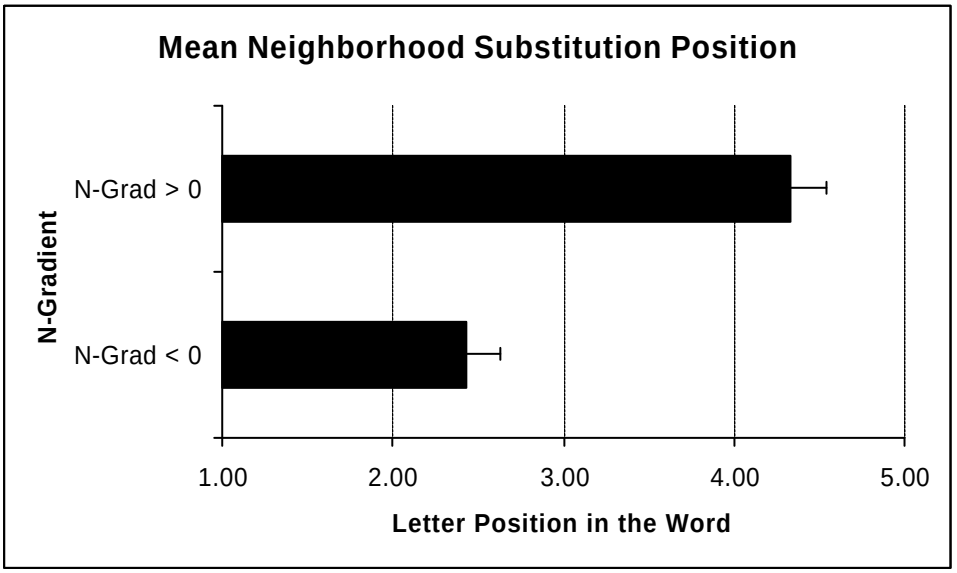


Figure 34: Mean neighborhood substitution letter position for positive and negative N-Grad. Words with positive N-Grad were substituted by positional neighbors on letter positions much more to the end of the word.

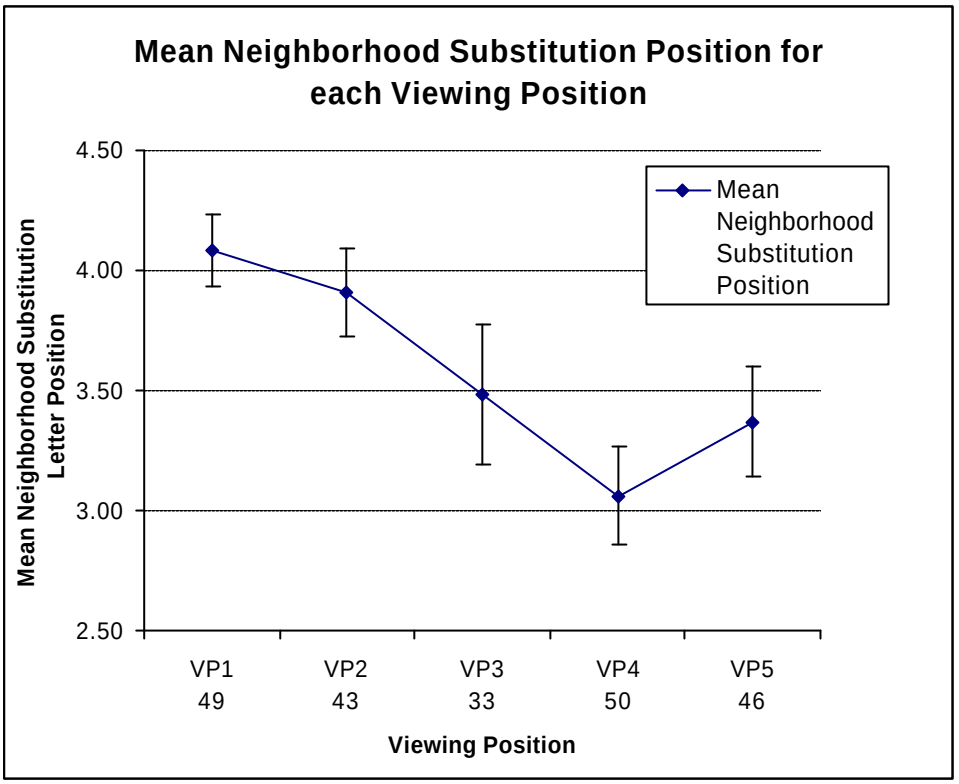


Figure 35: Mean neighborhood substitution letter positions for all five viewing positions. The more initial the viewing position was, the more final was the neighborhood substitution letter position. For viewing positions 1 to 4, there was an almost linear decrease in average neighborhood substitution position with increasing viewing position. As for the N-Grad effect for overall accuracy, position 5 was special, as mean neighborhood substitution position started to increase again. Below the abscissa, the number of neighborhood substitutions per viewing position is given. Note that although overall performance was worse on the final two viewing positions than on the initial two viewing positions, the number of neighborhood substitutions was almost equal. As each word was presented equally often on each viewing position, the results depicted in this figure cannot be due to different neighborhood distributions.

When I plotted the accuracy effect of the N-Grad against the average neighborhood error position for all five viewing positions, I observed an almost linear relationship (Figure 36). Note that these dependent variables are not naturally correlated, because the mean position of neighborhood substitutions – i.e., not their number which is a subset of errors – is investigated. Thus, there was something peculiar about the performance on viewing position 5 that occurred simultaneously in two independent analyses – in the overall accuracy analysis of the effect of the N-Grad manipulation and the average neighborhood error position. The relationship between these a priori independent variables was almost systematically linear for all five viewing positions (see Figure 36). However, while there was a monotonous negative linear effect for positions 1 to 4 for both variables, both effects were reversed in a similar manner for position 5.

So, for all positions including position 5, the relation of the two effects to each other was as expected, while the characteristics of the single effects themselves on position 5 were unexpected. I will elaborate on this finding in the discussion.

Following a similar analysis in Experiment 4, I also studied in more detail the interaction between frequency and viewing position. I correlated mean accuracy (i.e., averaged over frequency groups) with the frequency effect on each viewing position and each N-Grad word group. The correlations and linear trends generally went in the same direction as in Experiment 4, but did not reach significance. Average performance was negatively correlated with viewing position when comparing middle- and low-frequency words($r = -.12$, linear trend: $t(8) = 0.36$; $p > .10$) and for middle- and high-frequency words, if the unexpected drop-off for positive N-Grad on position 5 discussed above was eliminated ($r = -.18$, linear trend: $t(7) = 0.50$; $p > .10$; without elimination $r = .20$, linear trend: $t(8) = 0.60$; $p > .10$). In sum, as neither the variation of visibility nor the variation of performance was as large as in Experiment 4, it is not surprising that no interaction was found in Experiment 6. This analysis represents an example of how a low power of visual manipulation may produce a null effect which is not wholly conclusive.

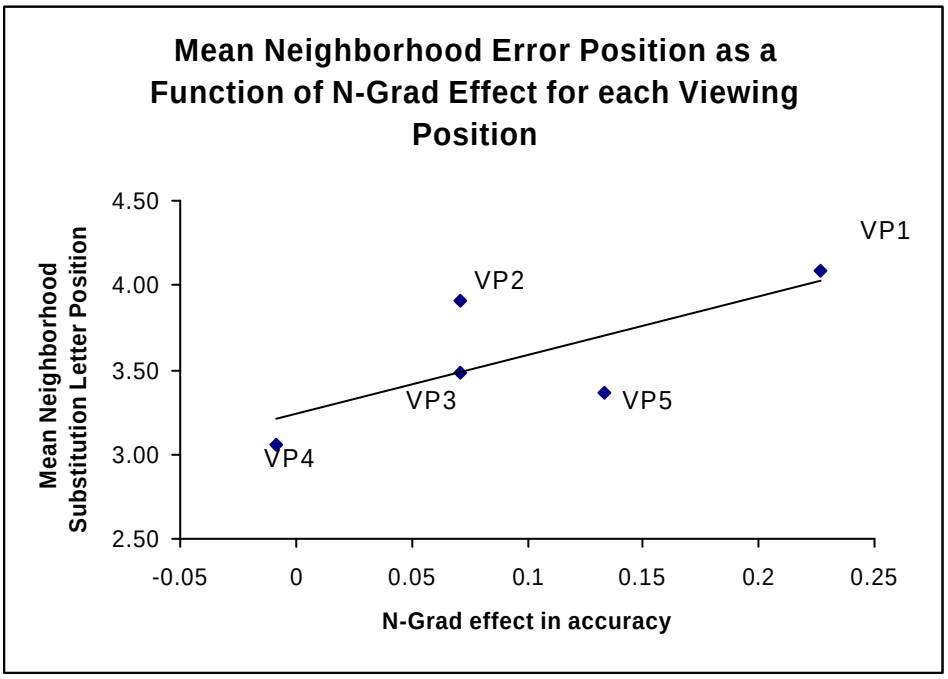


Figure 36: Relation between N-Grad effect (accuracy for negative N-Grad minus accuracy for positive N-Grad) and mean neighborhood substitution position over viewing position. As the N-Grad effect increased, mean neighborhood substitution position shifted more to the right. The relation was systematic, but not monotonic for viewing positions, as viewing position 5 was intermediate in both not naturally confounded measures.

Finally, raw analysis of neighborhood position errors over all viewing positions confirmed the interaction of the N-Grad manipulation with viewing position (see Figure 37). Although the same words were presented equally often on all viewing positions, notable differences could be observed for the distribution of neighborhood errors over the different viewing positions ($chi^2(24) = 107$, $p < .001$)[12]. Consider, e.g., the neighborhood positions 2 and 4 used by Grainger et al. (1992). There were not many neighborhood substitutions when viewing position equaled neighborhood position. However, in the crossed conditions (for viewing position 2 and neighborhood position 4, and for viewing position 4 and neighborhood position 2) there were much more neighborhood substitutions (see Figure 37). More generally, a systematic interaction neighborhood substitution position and viewing position could be observed. As regards viewing positions 1 and 2, more neighborhood errors could be observed for the final letter positions, while for viewing position 3 the number of errors increased at both ends of the word and was minimal for position 3. Note that neighborhood position 5 almost always produced the highest number of errors even when fixation was on that very position 5. This is in contrast to the results obtained for all other positions. While absolute accuracy was, as always in viewing position experiments, worst on viewing position 5, the absolute number of neighborhood substitutions as a subset of these errors was not highest, but only medium. Errors on viewing position 5 were quite often not neighborhood substitutions (look at the overall level of the curves in Figure 37).

The overall number of neighborhood errors on neighborhood position 5 was by far the highest and for position 1 the lowest. This result contrasts markedly with the overall number of positional neighbors in the current stimulus group. There were three times more neighbors on letter position 1 than on letter position 5 (see Figure 37). The overall number of neighborhood substitutions for each letter position did, thus, not correspond well to the respective number of positional neighbors on that letter position. In contrast, the two measures were negatively correlated ($r = -.32$). Thus, a neighborhood substitution on letter position 5 was overall more than ten times more likely than a neighborhood substitution on letter position 1 when all viewing positions are considered (for central fixation only the ratio is 5.89). This huge probability difference questions the computation of the N-measure which assumes that all neighbors on all positions are equally influential in orthographic space at least for PITs. Neighbors on the final position seemed much more influential in this experiment.

---

[12] Due to the Latin Square Design and the uneven distribution of neighbors over letter positions the single entries of the viewing position by neighborhood distribution matrix are not independent, and the prerequisites for any statistical test including the chi^2 test are hurt. The significance level may be somewhat affected by this violation. Practically, the effects are so huge ($chi^2(24) = 107$, see Figure 37) that these considerations have largely theoretical implications.

All these analyses seem to confirm the hypothesis that word recognition systematically varies as a function of the relationship between orthographic information distribution and viewing position and not only as a function of their main effects. This implies that neighborhood effects at least in a PIT cannot fully be understood as orthographic-lexical or phonological-lexical effects but rather as indexes of similarity in visual-orthographic-lexical space.

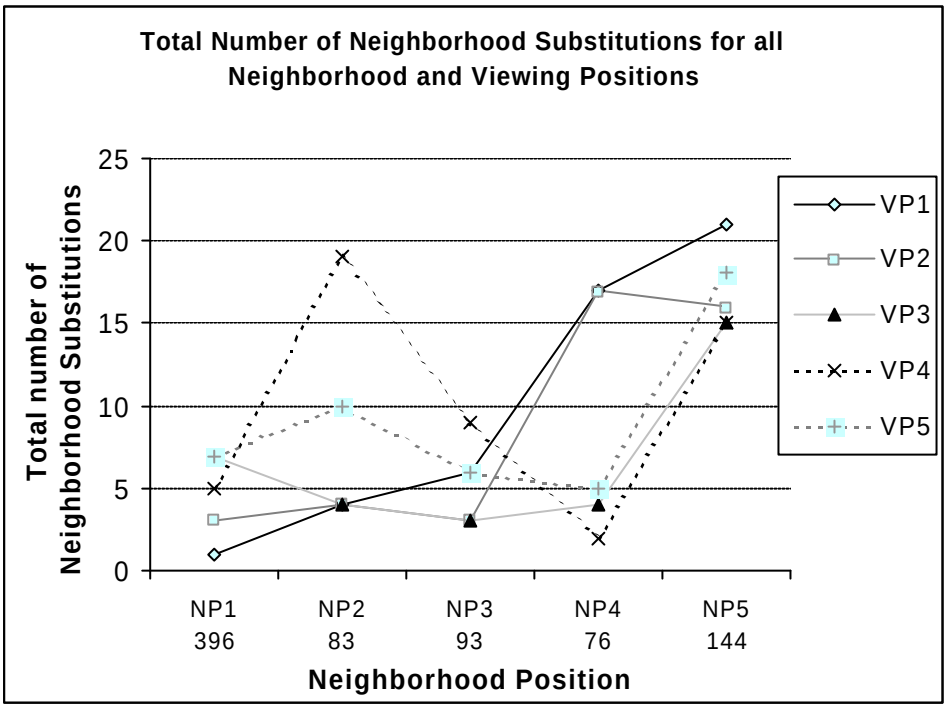


Figure 37: Number of neighborhood substitution for all five letter positions and all five viewing positions. Below the letter positions on the abscissa the total number of positional neighbors in the current stimulus sample is indicated. See, for example, the interaction between viewing position and neighborhood substitution position for viewing positions 2 and 4 and letter positions 2 and 4 (cf. Grainger et al., 1992). Note that although on letter position 1 much more neighbors existed than on letter position 5, less neighborhood substitutions were made on all viewing positions (see text).

2.4. Model Fits

The fits with the model of Nazir et al. (1998) also demonstrated that there was a shift in asymmetry as a function of the N-Grad. For a positive N-Grad, the asymmetry ratio between the drop-off rates to the left and to the right was close to 1 for all frequency groups and on average (see Table 12). Hence, the fitted curves were almost symmetrical. For negative N-Grad the asymmetry ratio was close to 2 for all frequency groups and on average (see Table 12). Hence, the fitted curves showed an asymmetry: Recognition was markedly better at the initial positions than at the final positions. Averaged empirical

viewing position curves for both N-Grads and the respective fitting curves are shown in Figure 38. The fit of the negative N-Grad curve was quite good (RMSD = .017) while the fit of the positive N-Grad was not as good (RMSD = .051). For positive N-Grad, the decrease in accuracy between positions 4 and 5 (see Figure 38) was underestimated by the model. This corroborates with the counterintuitive empirical results obtained at viewing position 5. While from viewing positions 1 to 4 the accuracy difference between words with positive and negative N-Grad had decreased with increasing viewing position, it had increased again on position 5 (see Figure 32). This unexpected drop-off in accuracy for the VPE curve for positive N-Grad was resembled by the mismatch on the model for the final two positions for this curve.

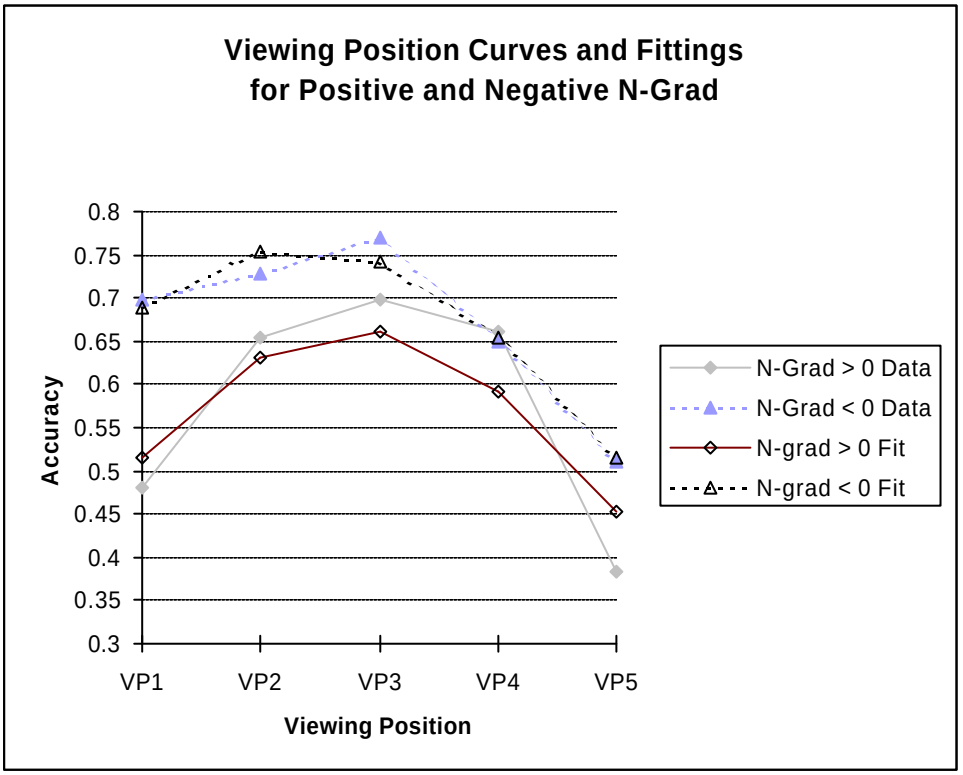


Figure 38: Data fits for the two N-Grad groups. Fits for positive N-Grad are much more symmetrical than fits for negative N-Grad. Again viewing position 5 is special. The fit for positive N-Grad is worse, due to the unexpectedly strong drop-off from position 4 to 5 in that condition.

Except for the final positions of the VPE curve for positive N-Grads, the model fitted the empirical data well. However, these fits could only be obtained by the model when adjusting the asymmetry parameter as a function of information distribution in the word. Hence, the asymmetry parameter in the current form of the model cannot be regarded as a purely visual parameter in this experiment. Similarly, the drop-off rates for the curves for different frequencies also varied. Thus, as also found in Experiment 4, the drop-off rate is influenced by lexical attributes of the presented stimuli.

Fitting Parameters and Fitted Values for All Viewing Positions									
N-Grad	**Frequency**	d	A	RMSD	VP1	VP2	VP3	VP4	VP5
N-Grad > 0	**all F**	0.060	1.18	0.051	0.52	0.63	0.66	0.59	0.45
N-Grad < 0	**all F**	0.035	1.71	0.017	0.69	0.75	0.74	0.65	0.51
N-Grad > 0	**low F**	0.080	1.17	0.034	0.40	0.53	0.57	0.49	0.33
N-Grad > 0	**middle F**	0.054	1.23	0.028	0.55	0.66	0.68	0.61	0.48
N-Grad > 0	**high F**	0.047	1.15	0.102	0.60	0.70	0.73	0.67	0.55
N-Grad < 0	**low F**	0.055	1.71	0.030	0.55	0.64	0.62	0.50	0.33
N-Grad < 0	**middle F**	0.028	1.78	0.021	0.75	0.80	0.79	0.71	0.59
N-Grad < 0	**high F**	0.025	1.71	0.023	0.77	0.82	0.81	0.75	0.64

Table 12: Fitting Parameters, deviations and fitted accuracy. RMSD: Root mean square deviation, d: Drop-off parameter to the right, A: Asymmetry parameter between drop-off to the right and drop-off to the left; the bigger A, the more the viewing position curve is shifted to the left, VP: Viewing positions within a word, N-Grad – neighborhood gradient (see text for computation), frequency (F) Groups: Low : F < 10, middle: 10 < F < 50, high: F > 50.

2.5. Discussion

Empirical results and model fits both indicate a visual alteration of orthographic neighborhood size effect because positional neighborhood distribution interacted with viewing position. Therefore, neighborhood size, at least in this PIT, cannot be regarded as an orthographic-lexical effect but rather as a visual-orthographic-lexical effect. Particularly on the initial viewing position, participants profited when the positions of information ambiguity were also on the initial letter positions. This finding was consistent for all frequency groups. While words with negative N-Grad were better recognized on the initial viewing positions than words with positive N-Grad, this was not the case for the final viewing positions. Error analyses seemed to confirm this result. The more initial the fixation position was in the presented word, the more final were confusion positions of erroneously reported neighbors. The only notable exception in both analyses was viewing position 5. A possible explanation is that legibility of the word is so bad on the final viewing position that often many letters are not recognized. Then neighbors that differ by only one letter from the target would only be a small part of the possible lexical candidates for recognition. This hypothesis is confirmed by the fact that there are relatively few neighbor substitutions on position 5

compared to the overall number of errors made on that position (see Figure 37). However, this hypothesis and the reliability of this diverging result needs further investigation in the future.

The model fits were also straightforward. Negative N-Grad increased the asymmetry ratio and made the fitting curve more asymmetrical, while positive N-Grad words produced more symmetrical fitting curves. The fits confirmed that the results obtained in the ANOVA were specific in that N-Grad changed asymmetry in the specific hypothesized direction. Furthermore, they demonstrate that the asymmetry parameter in the current model is not a purely visual parameter.

To investigate the interaction between N-Grad and viewing position I chose words with rather steep N-Grad for all six stimulus groups. It was not possible to match these groups on all other attributes; therefore the main effect of the N-Grad manipulation needs careful interpretation. Grainger et al. (1992) have interpreted their results as possibly indicating sequential bias left to right processing by imposing more facilitatory feedback from the initial position compared to the final positions. “In this case, increasing the visibility of the initial letters by [fixation on the initial letters] would increase the activation levels of initial letters, thus causing a relatively greater increase in activation at the word level than when end letters are rendered more visible.” While this interpretation only accounts for the main VPE, the interaction of viewing position with neighborhood gradient is also consistent with their interpretation. *“If word-initial letters produce a greater rise in activation of lexical representations than do word-final letters, it is conceivable that the difference in activation level of a competing pair [with ambiguity on one of the initial positions] is less than the difference in activation level of a competing pair [with ambiguity on the final letter positions].”* (p. 55). Given that a lesser difference to neighbors as lexical competitors produces greater inhibition, this interpretation could account for the interaction between viewing position and neighborhood gradient as well as for the main N-Grad effect.

In sum, as a fixation position manipulation has been shown to produce huge effects in all kinds of word recognition tasks (cf. O’Regan & Jacobs, 1992), these interactions of viewing position with neighborhood distribution question the generalizeability of all neighborhood effects obtained at central viewing position with the principle of isolated variation. However, this result is yet specific to the visual manipulation of viewing position. To draw more general conclusions, it is necessary to examine whether visual attributes do *generally* influence the pattern of neighborhood effects in their visually determined orthographic-lexical hyperspace. This will be done Experiment 7.

3. Experiment 7: A Visual Neighborhood Confusability Effect

3.1. Introduction

Experiment 7 was designed to generalize the results of Experiment 6 to gain independent evidence for the hypothesis that neighborhood is better conceptualized as a similarity index of visual-orthographic-lexical space than a reduced orthographic-lexical space only. A couple of objections could be raised against Experiment 6. First, it was not possible to hold syllable number constant for German five-letter lemmas. The manipulation of the neighborhood gradient was chosen such that opposite neighborhood distributions were strongly biased towards either the end or the beginning of the word to avoid null effects due to a weak manipulation of the variable of interest. When, additionally to this manipulation, length was held constant across a frequency by N-Grad 3×2 design, it was not possible to still keep syllable number constant over the respective frequency groups. Although I am not aware of any interaction between syllable number and viewing position, it is certainly more satisfactory to exclude this confound as an alternative account in another experiment. Second, it was argued against PITs that accuracy as dependent variable may reflect sophisticated guessing strategies that could obscure the lexical retrieval process (Andrews, 1997; Massaro, Taylor, Venezky, Jastrzembki, & Lucas, 1980; Paap & Johansen, 1994). Therefore, it would be helpful to employ a PIT, in which accuracy is not the only dependent variable in order to disentangle the confound between task and dependent variable (see also Ziegler et al., 1998). The word fragmentation task (see Experiment 5 and Jesse, Nuerk, Graf, & Massaro, 1999; Nuerk, Graf, Boecker, Gauggel, & Jacobs, in press; Snodgrass & Mintzer, 1993; Snodgrass & Poster, 1992; Ziegler et al., 1998) offers such a possibility. Recall from the introduction to Experiment 5 that Ziegler et al. (1998) computed a regression analysis in which both dependent variables, accuracy and level of correct recognition, were sensitive to letter confusability. However, as concerns N, the effects differed between the two dependent variables. The level (i.e., speed of response) was sensitive to log frequency and - in a partial correlation - to bigram frequency, but not to N. In contrast, accuracy was shown to be sensitive to N, HFN and positional letter frequency (cf. Grainger & Jacobs, 1993); however, when a partial correlation was computed the HFN effect disappeared. Snodgrass and Mintzer (1993) obtained diverging results. They found inhibitory effects of neighborhood in both measures, in accuracy and in level of recognition, when no sophisticated guesses and hypothesis testing were possible on earlier levels (Experiments 3 – 5).

A possible account for these diverging effects of different studies is that the word fragmentation task is sensitive to N as an index of similarity in visual-orthographic-lexical space. While N as an index of orthographic-lexical space was similarly manipulated in the two experiments, the targets in the two studies might have differed with respect to the visual similarity / confusability of their competing neighbors. However, this alternative account lacks empirical support. It has never been shown that visual properties of N could alter its effects in a word fragmentation task. This shall be done in Experiment 7.

How can one investigate neighborhood as a visual-orthographic-lexical measure in word fragmentation tasks? While in Experiment 6 I varied letter legibility indirectly by varying viewing position, I directly manipulated visual letter confusability in Experiment 7. I used a letter confusability matrix to investigate how likely each letter is confused with each other letter in that very fragmentation task on the low recognition level 2 (see Ziegler et al., 1998; and Appendix H). Then, I computed the neighborhood confusability index N-conf for each German five-letter word in the fragmentation task simply by summing up the relative frequencies of the critical letter exchanges disambiguating a target word from its neighbors.[13] Consider, for example, the word “PEGEL”. It has the highly confusable neighbor “REGEL”, which is created by an exchange of letter P with letter R. The relative frequency of this exchange is 0.10 (see Appendix H). By summing up the frequencies for all neighborhood exchanges, I computed a neighborhood confusability index N-conf. If neighborhood size is fully understood as a non-visual orthographic-lexical effect, visual manipulation of letter confusability of the critical letters should not influence performance. If neighborhood size is better understood as a visual-orthographic-lexical effect, words with highly confusable neighbors (high N-conf words) should at least produce more errors and, specifically, more erroneous neighborhood responses. If this effect is due to sophisticated guessing strategies of lexical retrieval in a serial model, I should observe this effect for accuracy, but not speed of recognition. However, if high N-conf words are also slower, then this would in my opinion indicate a lexical competition and inhibition that slows down responses for these words and does not simply open up more and more likely guessing opportunities. Consequently, slower responses for highly confusable neighbors would then be hard to reconcile with serial stage models.

An additional more specific question of Experiment 7 was whether neighbors do have inhibitory effects in this PIT even if they are not highly visually confusable. Therefore I compared low N-conf words to

[13] Words with “ß” or Umlaut (i.e., Ä, Ö, Ü) were not included in the N-conf computation as for those words no letter confusability measure existed (see Appendix H). Participants were informed that such words were not part of the experiment.

hermits, i.e., words with N = 0. I wanted to examine the question of whether visual confusability of the critical letters is a necessary prerequisite for neighbors to be inhibitory or not. If it is necessary, then only high N-conf words should be affected by neighborhood inhibition. If it is not necessary, low N-conf words should also be affected by inhibition when performance is compared with that of hermits. Can inhibition in orthographic-lexical space be demonstrated when the lexical competitors are almost unconfusable?

As for the comparison between high and low N-conf words, again speed of recognition and accuracy can be investigated for the comparison between hermits and low N-conf words. If inhibitory neighborhood effects are due to sophisticated guessing, I should only observe them in the accuracy measure, however, if those effects are also due to lexical competition and inhibition I should also observe them in the speed measure. However, the opposite effect could also be true. Facilitatory effects of neighborhood size in the NT and the LDT are normally attributed to their more frequent orthographic and phonological sublexical structure (Andrews, 1997; Forster & Taft, 1994; Ziegler & Perry, 1998). With visual-orthographic-lexical interference being low I might actually observe a facilitatory effect of neighborhood size in this PIT (Grainger et al., 1989) that has to my knowledge not been reported in PITs previously.

In sum, the fragmentation task allows a very fine-grained analysis of visual-orthographic-lexical effects because two different dependent variables in one experiment can be examined. It diminishes the risk of obtaining null effects, because it has been shown to be highly sensitive to letter confusability, which is the base of the visual manipulation used to separate stimulus groups with different neighborhood confusabilities.

3.2. Method

3.2.1. Participants

11 psychology students of the Philipps-University Marburg participated in the experiment. All were native German speakers and had normal or corrected to normal vision. They received course credit for participation in the experiment. None of them had participated in the other experiments.

3.2.2. Stimuli and Design

120 German lemmas from the CELEX- Database (Baayen, et al., 1993) were used in Experiment 7. All stimuli were five-letter German bisyllabic words (see Appendix I). For the present experiment, three groups were important. One group were hermits, one group had low confusable neighbors as indicated by a low index N-conf (N-conf < .05; see introduction to Experiment 7), and a third group with same N but

high N-conf (N-conf >= .10). Those three groups were matched for overall letter confusability, which was computed in the same way as in Experiment 5. Thus, an effect of neighborhood confusability cannot be attributed to overall letter confusability, but specifically to confusability of the critical disambiguating letters of a target and its respective neighbors. Moreover, N and HFN were matched between neighborhood groups (see Table 13 for stimulus characteristics). Stimuli in all three experimental groups consisted of high-confusable letters. Hence, in the experiment letters with high confusability indices would have been very frequent, if only those three groups had been examined. Therefore, a fourth group of stimuli (hermits) with low confusability indices was added to the experiment to reduce a possible experimental bias towards certain more frequent letters that could mask other effects. Additionally, this group provided an additional control comparison with the highly confusable hermit group.

	log Freq	Syllables	N	N-conf	HFN	LCI
High N-conf Words	0.49	2	4.90	0.19	2.57	0.68
Low N-conf Words	0.49	2	5.03	0.02	2.57	0.67
Hermits	0.48	2	0.00	0	0.00	0.69
Controls	0.48	2	0.00	0	0.00	0.37

Table 13: Stimulus characteristics in Experiment 7: log Freq: Mean logarithmic (frequency per million + 1) to the base 10, Syllables: Number of syllables, N: Number of neighbors, HFN: Number of higher frequency neighbors, N-conf: Neighborhood confusability, LCI: Letter confusability index.

3.2.3. Procedure

The screen fragmentation procedure was identical to that of Ziegler et al. (1998) used in Experiment 5. An experimental session was subdivided into two blocks of 60 words in random order for each participant. Prior to the experiment participants received eight trials to familiarize them with the experiment.

3.3. Results

The main result of Experiment 7 was that low N-conf words were faster and more accurately recognized than high N-conf words (accuracy: $t_1(10) = 4.29$; $p < .001$; $t_2(58) = 2.79$; $p < .001$; recognition level: $t_1(10) = 2.33$; $p < .05$; $t_2(58) = 1.79$; $p < .05$, see Figures 39 and 40). Thus, neighborhood confusability had a clear inhibitory effect that was not mediated by any SATO. This result could also convincingly be observed for the single participants (see Figure 42). High N-conf words were also less accurately recognized and tended to be recognized more slowly than the respective hermits (accuracy: $t_1(10) = 4.29$; $p < .001$; $t_2(58) = 3.75$; $p < .001$; recognition level: $t_1(10) = 2.38$; $p < .05$; $t_2(58) = 1.01$; p =

.16, see Figures 39 and 40). In contrast, low N-conf words were neither significantly slower nor significantly less accurately recognized than hermits, rather, descriptively they even tended to be faster, but less accurate (accuracy: $t_1(10) = 1.49$; $p = .08$; $t_2(58) = 1.15$; $p = .13$; recognition level: $t_1(10) = .78$; $p = .23$; $t_2(58) = 0.82$; $p = .21$). Because Snodgrass and Mintzer (1993) had also suggested directly analyzing whether really similar alternative lexical entries are retrieved, I looked at neighborhood substitutions as in Experiment 6 (see Figure 41). The results were straightforward: High N-conf words were more often confused with their neighbors than low N-conf words ($t_1(10) = 3.54$; $p < .01$; $t_2(58) = 1.64$; $p = .05$), but even low N-conf words still produced significantly more neighborhood exchanges than hermits, i.e., significantly more neighborhood substitutions than 0 ($t_1(10) = 3.54$; $p < .01$; $t_2(58) = 3.79$; $p < .001$). Obviously, high N-conf words were also more often confused with neighbors than hermits ($t_1(10) = 3.54$; $p < .01$; $t_2(58) = 6.70$; $p < .001$; see Figure 41). All these calculations were performed for hermits with a similar LCI. In the control condition, hermits with low letter confusability were faster and more accurately recognized than hermits with high letter confusability (accuracy: $t_1(10) = 0.19$; $p = .43$; $t_2(58)= .25$; $p = .40$; recognition level: $t_1(10) = 8.54$; $p < .001$; $t_2(58) = 5.00$; $p < .001$), thus validating the letter confusability manipulation. This control condition is important because the obtained pattern of results implies that letter confusability effects in the fragmentation task (cf. Experiment 5) cannot be totally driven by neighborhood confusability.

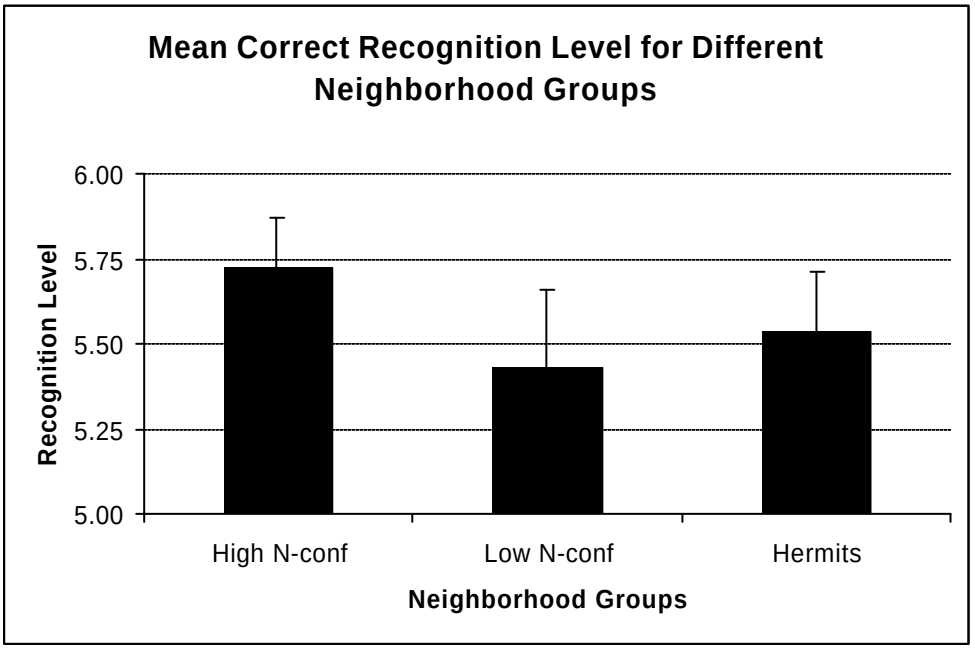


Figure 39: Mean recognition level in the fragmentation task. High N-conf words were recognized slower than low N-conf words and hermits. Low N-conf words were not significantly slower than hermits (and descriptively even faster).

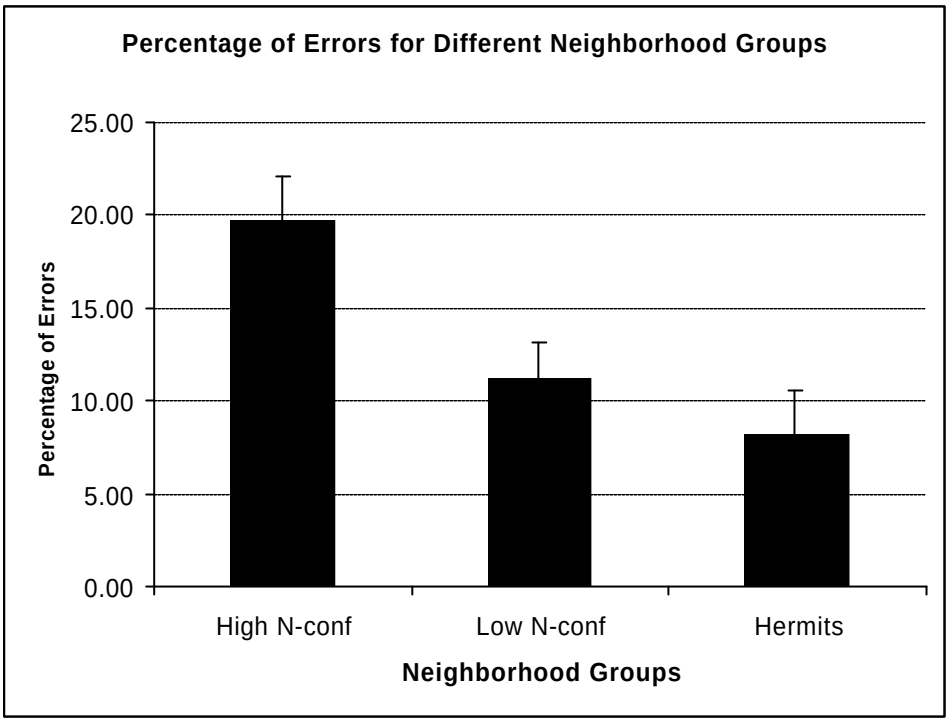


Figure 40: Mean percentage of errors in the fragmentation task. High N-conf words were less accurately recognized than both, low N-conf words and hermits. In contrast, low N-conf words did not differ significantly from hermits.

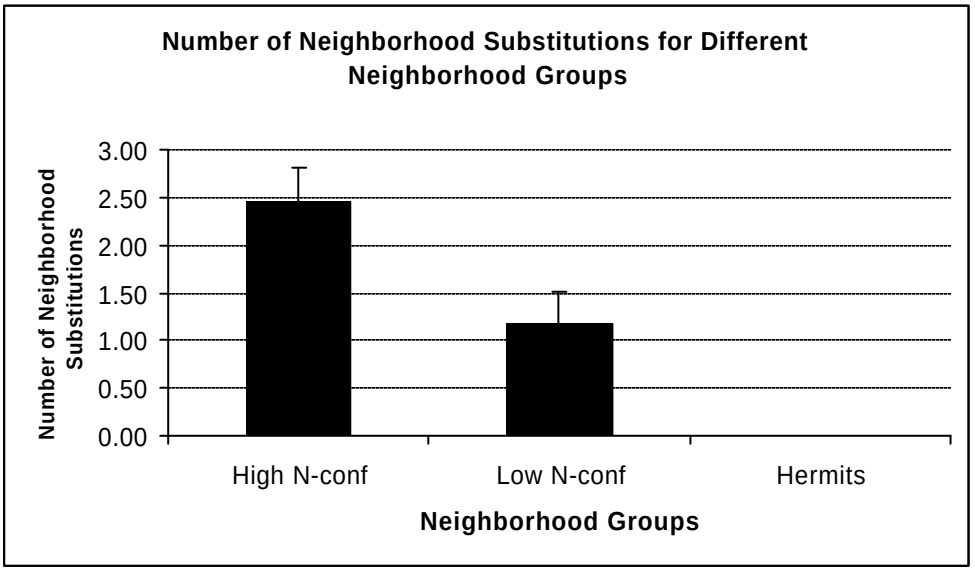


Figure 41: Mean number of neighborhood substitutions. High N-conf words were more often confused with their respective neighbors than low N-conf words and hermits, and even low N-conf words were more often confused with neighbors than hermits.

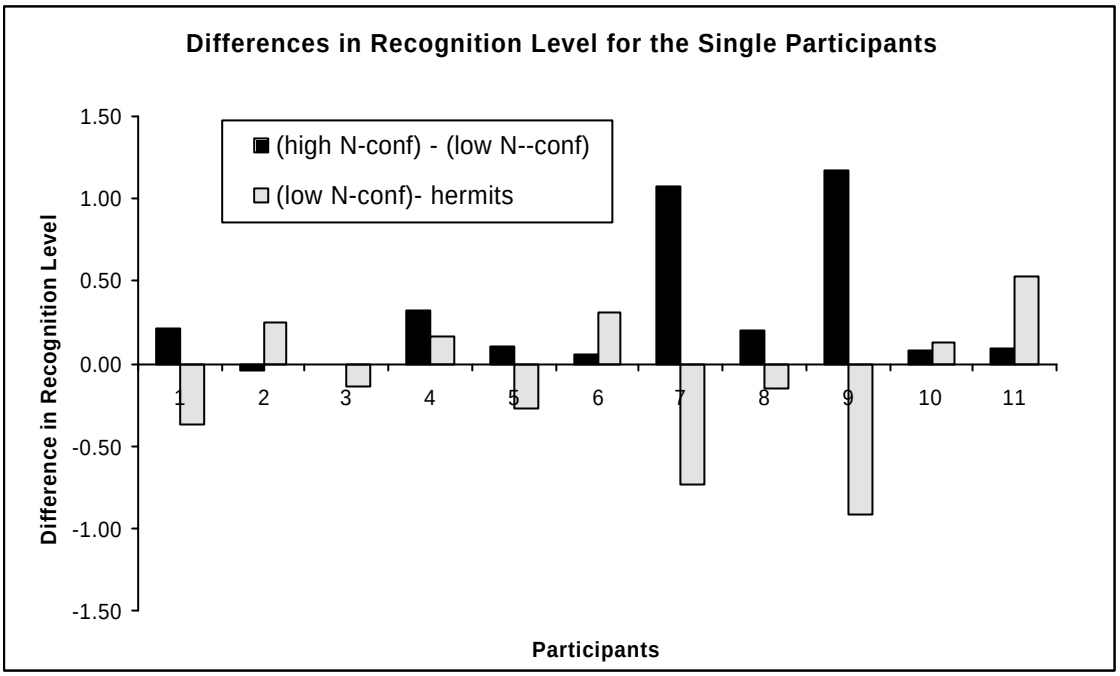


Figure 42: Differences in Recognition Level for the single participants. Even for the less pronounced recognition level effect, nine out of 11 participants tended to be descriptively slower for high N-conf words than for low N-conf words. In contrast, low N-conf words are not consistently slower than hermits, instead, in six of 11 participants, hermits were faster.

3.4. Discussion

Experiment 7 confirmed the hypothesis that at least in PITs neighborhood size effects must be understood as visual-orthographic-lexical effects rather than orthographic-lexical effects. High N-conf words were slower and less accurately recognized than low N-conf words and they were also more often confused with their neighbors. While this visual-orthographic effect was highly significant for all dependent variables, the comparison between low N-conf words and hermits revealed no significant results, neither for accuracy nor for recognition level as dependent variable. Thus, orthographic ambiguity had no inhibitory effect in the absence of visual ambiguity. That low N-conf words did not differ in accuracy is remarkable, since even low N-conf words were significantly more often exchanged with neighbors than hermits. As those exchanges constitute a subgroup of all errors analyzed in the accuracy analysis, it is surprising that the low N-conf words did not differ in overall accuracy. In this study, overall accuracy as a dependent variable does not seem to index the same processes as neighborhood exchanges.

How do my results correspond with Ziegler et al.'s (1998) results where N was an inhibitory predictor for accuracy, but not for recognition speed? Ziegler et al. (1998) did not distinguish between high and low neighborhood confusability. An explorative data analysis revealed that if I also did not distinguish between high and low confusable neighbors, but simply put together all words with neighbors and compared them

to hermits, I obtained the same pattern of results as Ziegler et al. (1998) in their regression study. Words with neighbors were responded to less accurately than hermits, but not slower (accuracy: $t_1(10) = 4.56$; $p < .001$; $t_2(88) = 2.76$; $p < .01$; recognition level: $t_1(10) = 0.48$; $p > .10$; $t_2(88) = 0.08$; $p > .10$). Thus, the apparent confound between neighborhood effects and dependent variable in the fragmentation task was only obtained when neighborhood confusability was not controlled. However, the question remains of why such a clear dissociation between overall neighborhood effects and dependent variable was obtained in the overall analysis, but not in the inference-statistical analyses for the different neighborhood confusability groups. The answer is that information is always lost in inference-statistical conclusions. The difference between the dependent variables was also present in the descriptive analysis of the single neighborhood confusability groups, but it did not reach statistical significance for the single groups. Low N-conf words descriptively tended to be faster than hermits but less accurate, while for high N-conf words the accuracy results were much clearer than the results for recognition level.

This pattern of results points to a possible confound between task and dependent variable. N effects in the PIT that are mostly found to be inhibitory (cf. Andrews, 1997) may rather be facilitatory for accuracy as a dependent variable, but not as much for recognition speed. In the framework of interactive activation models (Jacobs et al., 1998), the results may indicate that particularly for low N-conf words their overall activation may not be lower than that of hermits. Instead, the activation of their neighbor competitors may also be so high that confusions with those neighbors may occur, which may then lead to lower accuracy. This converges with the observation that despite a significant accuracy effect, still significantly more neighborhood exchanges were observed for low N-conf words. That lowly confusable neighbors were not slower than hermits, however, does not argue against reciprocal inhibitions at the lexical level (cf. Grainger & Jacobs, 1996, Andrews, 1997 for discussion), because neighbors tend to have a more frequent sublexical structure than hermits (e.g., Forster & Taft, 1994; Nuerk, Rey, et al., 2000; Ziegler & Perry, 1998). According to most accounts a more frequent sublexical structure tends to be facilitatory, while reciprocal inhibition on the lexical level tends to be inhibitory. These processes seem to cancel each other out in the fragmentation task for low N-conf words with a tendency for the facilitatory process being more prevalent in recognition speed and for the inhibitory process being more prevalent in accuracy results. This idea is also supported by the results for hermits with high and low overall letter confusability. Because for both groups no lexical neighborhood competitors can possibly be retrieved, no inhibitory lexical processes should exert their influence on accuracy as a dependent variable. This was indeed the case; an accuracy effect between these groups was totally absent while clear effects for the recognition

level were observed. Thus, all these considerations point towards the idea that lexical competition processes indeed seem to affect the dependent variable accuracy more, while a more frequent and less confusable sublexical structure may particularly affect recognition speed.

Although accuracy seems to be more affected by lexical competition processes than recognition speed, a simplified relation of accuracy data indexing lexical competition and recognition speed data indexing visual or sublexical processes is not consistent with the results. High N-conf words were still not recognized as well as low N-conf words in all dependent variables. Because there is no difference between high and low N-conf words in the number of neighbor competitors nor in the overall visual letter confusability, the results clearly suggest that both variables are sensitive to visual-orthographic-lexical interactions. Finally, these differences in neighborhood confusability may offer an explanation for the diverging results of Snodgrass and Mintzer (1993), who found neighborhood inhibition for the recognition level, too, and the results of Ziegler et al. (1998), who did not. Thus, results obtained with one dependent variable may not be obtained with another dependent variable. Consequently, future studies should investigate more than one dependent variable.

4. Discussion of Study 3

The results of Experiments 6 and 7 support the idea that, at least in PITs, neighborhood effects are better understood as visual-orthographic-lexical effects rather than purely orthographic-lexical effects. The results are a generalization of similar findings of Grainger et al. (1992). In particular, these results show that the visual modulation of the neighborhood effects is not only observed when a critical disambiguating letter is on or off fixation. Instead, in my study, neighborhood size and visual determinants interacted in a characteristic way. Neighbors on letter positions that were far away from fixation and thus not well legible tended be particularly detrimental for performance in a PIT. In a similar way, disambiguating letters of two neighbors seemed to be particularly inhibitory if they were visually highly confusable. Finally, the results of these studies also indicate that reviews (Andrews, 1997) or models (e.g., Plaut et al., 1996) that do not consider visual attributes of neighborhood may fall short of generally accounting for neighborhood effects. These models may only do so under very restricted near-optimal presentation conditions.

I found strong visual-orthographic lexical interactions in two PITs. In Experiment 6 viewing position generally interacted with neighborhood position in a specific hypothesized way. If a word possessed many neighbors at the beginning of the word, fixation at the beginning of the word improved performance

relatively to words with neighbors at the end. The more fixation moved toward the end of the word, the more this disadvantage of having neighbors at the end of the word disappeared except for position 5.[14] However, for all other letter positions and for all different viewing position curves the interaction was as expected. Neighborhood at the beginning of a word produced much flatter viewing position curves and model fits showed that the asymmetry parameter of the model of Nazir et al. (1998) systematically varied as a function of neighborhood distribution. Thus, the asymmetry of the viewing position curve was not purely influenced by perceptual learning due to reading directions (Nazir, 2000), but the VPE was influenced by the orthographic information distribution within the word (see also Beauvillain et al., 1996; Brysbaert et al., 1996; O'Regan et al., 1984; Underwood et al., 1990 for similar results not investigating orthographic neighbors). However, as this interaction was asymmetrical - the viewing position curve became symmetrical for words with disambiguating information at the end, but heavily asymmetrical for words with disambiguating information at the beginning – the results may question generalizations based on results obtained for central viewing position.

A similar argument holds for Experiment 7 where the confusability of the disambiguating letter of the target word with the respective letter of an orthographic neighbor was manipulated rather than its general legibility as in Experiment 6. High neighborhood confusability produced strong inhibitory effects in all dependent variables compared to hermits and to low neighborhood confusability. High N-conf words were slower, less accurate, and more often confused with neighbors than both low N-conf words and hermits. In contrast, low N-conf words did not significantly differ from hermits in either variable. Thus, null effects of N observed in PITs may be due to low neighborhood confusability and inhibitory effects may be due to high neighborhood confusability (cf. Andrews, 1997; for the fragmentation task: Snodgrass & Mintzer, 1993; Ziegler et al., 1998). The null overall effect of neighborhood in recognition speed in contrast to the inhibitory effect in accuracy may also indicate a possible confound between task and dependent variable. Therefore, more than one dependent variable should be examined before a generalization on variable-independent word recognition performance is made. Moreover, neighborhood confusability should be controlled, because it does strongly mediate this possible confound between the dependent variable and

[14] Only for position 5 did the results point in an unexpected direction: Performance for words with neighbors at the final positions was worse contrary to expectations – a possible explanation for this effect is that overall recognition at position 5 was so bad, and often so few letters were recognized that neighborhood position does not play such a role as for recognition on other positions. This account is supported by the literature and by the neighborhood position analysis: Performance on the final position is known to be by far the worst in all kinds of tasks (Nazir et al., 1998; O'Regan & Jacobs, 1992; O'Regan et al., 1984). In addition, while neighborhood confusion position shifted systematically more to the left with a more final viewing position, for position 5 this effect disappeared.

the obtained pattern of results. Finally, Experiment 7 indicates that the general claim of study 3 that neighborhood is best understood as a visual-orthographic-lexical rather than an orthographic-lexical effect holds for different visual manipulations and stimulus groups.

In sum, the two experiments supported the notion that visual bottom-up processes such as letter legibility or letter confusability (Nazir et al., 1998; Ziegler et al., 1998) and orthographic-lexical processes such as competition in lexical retrieval (Andrews, 1989, 1997) cannot be separated, but do fundamentally interact. Therefore, visual attributes of neighbors should be controlled in experimental manipulations and taken into account in reviews. Furthermore, visual attributes should be incorporated in models of visual word recognition, if a given model should account for neighborhood effects and their underlying processes (as in Coltheart, Curtis, Atkins, & Haller, 1993; Jacobs et al., 1998; McClelland & Rumelhart, 1981; Paap et al., 1982; Rastle & Coltheart, 1999, but see Plaut et al., 1996), because neighborhood effects change as a function of presentation condition. If a manipulation of visual variables is not possible in a model, this model only valid for one visual presentation condition, which is mostly central fixation with undistorted stimuli under almost optimal illumination. However, this condition is not always found in natural reading, but is specific to single word presentation in laboratory experiments. In sum, this implies that measuring organization of knowledge of single words, be it a lexicon (Coltheart et al., 1993; Jacobs et al., 1998) or distributed representations (Plaut et al., 1998) depends not only on orthography, phonology, lexicality or semantics, but also on the simple visual characteristics interacting in a complex, non-linear manner with any of these variables.

With regard to the model of Nazir et al. (1998) and its refinement these results imply that the asymmetry parameter cannot be indexed as a purely visual parameter, because the asymmetry parameter is influenced by lexical information ambiguity. A refined model should take into account local information ambiguity as measured by neighborhood distributions as an additional source of asymmetry modulation in the viewing position curve if it is supposed to generally fit neighborhood effects.

V. General Discussion

1. The Perceptual Frequency Hypothesis

1.1. Introduction

This thesis provided evidence for the perceptual frequency hypothesis in two respects, that of item-specific visual word form (Study 1) and that of location (Study 2). Study 1 demonstrated in two PITs and one LDT that a word was better recognized in that particular item-specific visual form in which it was previously more frequently perceived. Study 2 provided evidence for the perceptual frequency hypothesis of location, as a word of a particular word length tended to be perceived best on those fixation positions on which a word of that length was usually most often encountered. Taken together, these findings suggest that elementary perceptual experiences, such as where and how a word has been seen in the past, have an influence on the reading process. In the next paragraph, I discuss the pattern of results for the two different aspects of the perceptual frequency hypothesis in greater detail.

1.2. The Perceptual Frequency Hypothesis of Item-Specific Word Form

The results of Study 1 provided evidence for the perceptual frequency hypothesis of item-specific word form: The more often a word was previously perceived in an item-specific visual word form, the better it was recognized in that particular visual form. German nouns, usually encountered with initial letter capitalization, were best recognized in that particular visual form, in both task environments. In the PIT, IUC form produced better performance than both AUC and ALC forms, and in the LDT the IUC form was superior to the ALC form. In sharp contrast to German nouns, a different pattern of results was obtained for German non-nouns which can be encountered in ALC form, but are printed in IUC form when they are functionally used as a noun (e.g., “Das Schwimmen macht Spaß.“ <Swimming is fun>) or when they are at the beginning of a sentence. As concerns non-nouns, IUC and ALC forms did not yield different results in the PIT. In the LDT, the pattern was even reversed for high-frequency non-nouns: ALC form led to faster responses than IUC form. Also, the comparison between all lower-case and all upper-case presentation modes differed between syntactic classes in the PIT. Non-nouns behaved like common English words (Paap et al., 1984) being significantly better recognized in lower-case than in upper-case, while for nouns the difference descriptively was in the opposite direction. In sum, the results are consistent with the perceptual frequency hypothesis of item-specific word form: Under equal experimental

conditions, nouns and non-nouns are differentially recognized better in their most frequent perceptual word form(s).

What do these results imply for alternative general accounts of case-mixing effects? In my opinion, any general account fails to explain the observed pattern of results: Abstract transletter or multiletter feature processing (Besner & Johnston, 1989; Mayall et al., 1997) should be similar for all Roman scripts and should not differ between nouns and non-nouns. Also, the idea, that I selected transletter features for nouns that were particularly recognizable in IUC form and transletter features for non-nouns that were particularly or equally well recognizable in ALC form could be rejected: No effects for nonwords with the same transletter features were observed. Abstract single letter processing (Evett & Humphreys, 1981; Paap et al., 1984; Rayner et al., 1980) also fails to provide a sufficient explanation for the fact that IUC and ALC forms consistently interacted with syntactic class in this study: An exclusive abstract letter processing account would predict no difference . Moreover, the distinctiveness and confusability of letter features (Paap et al., 1982; Paap et al., 1984; Rumelhart & McClelland, 1982, Rumelhart & Siple, 1974; Smith, 1969; Ziegler et al., 1998) does also not provide an explanation for the observed difference between nouns and non-nouns. If only individual confusability or distinctiveness of the initial letter were responsible for the diverging results concerning nouns and non-nouns the form of the viewing position curves should have changed in a characteristic way (for details see discussion of the curve fits in Experiments 1 and 2). Additionally, similar results as for words should have been observed for the corresponding nonwords in Experiment 3. Letter distinctiveness has previously been suggested as an account for the finding that normally English words are better recognized in lower-case than in upper-case (Paap et al., 1984). However, this account is not convergent with the current pattern of results. As for English words, I found a superiority in recognition performance for non-nouns which are - like English words - frequently encountered in ALC form. However, for words – that are almost never presented in ALC form – I failed to find such an effect. Rather, the difference was in the opposite direction and statistically syntactic class and word form (ALC and AUC) interacted between Experiments 1 and 2. Thus, my preferred interpretation of lower-case superiority would be a perceptual frequency account: Words are recognized best in the perceptual form they are most often encountered in. In English and for German non-nouns this form is the ALC rather than the AUC form. Finally, general word form accounts explaining mixed-case results by case-specific tuning, case swapping and lateral inhibition also do not provide a sufficient explanation for dissociating results for nouns and non-nouns.

Other possible alternative explanations of the present effects could also be dismissed. Language-specific accounts of German nouns being better recognized in IUC form cannot explain why a different pattern of results was observed for non-nouns. Task-specific accounts that question the generality of the results fail because differential word form effects were found in both, the LDT and two PITs. Strategic adjustments to blocked (or to mixed) presentation of syntactic classes with specific visual word forms also provide no explanation, because a dissociation of word form recognition with syntactic class was found in either presentation mode. Moreover, the perceptual frequency effect is also hardly a post-perceptual guessing effect. First, it was found for hermits, too, were guessing is –according to most models – not as ambiguos as for words with neighbors. Second, the effect was also found in a LDT. Finally, methodological and statistical objections that could be raised with regards to Experiments 1 and 2 in which the stimuli were repeated and the Latin Square design made a meaningful item-based analysis impossible were tackled in Experiment 3. Direct item-based comparisons in the absence of repetition effects provided the same pattern of inference-statistical results as participant-based comparisons. In sum, neither alternative general accounts of word form effects, nor language-, task-, or method-specificity can explain the pattern of results in all three experiments.

The claim that item-specific word form plays a role in the reading process seems to generalize from single word presentation to normal text reading as a recent study of Nißlein, Müsseler, and Koriat (1999) seems to indicate. Nißlein et al. (1999) investigated the missing letter effect (MLE) in German text reading. The missing letter effect first observed by Corcoran (1966) and termed as such by Healy (1976) is based on the observation that when readers are charged with detecting the letter t in continuous text they miss the letter more often when it appears as part of the word „the“ than when it appears as part of a less frequently occurring word. This missing letter effect has quickly been established as a very robust and reliable effect for other frequent function words like „of“ and „for“ also and moreover, it has been found to be translingually valid (e.g., Drewnowski & Healy, 1977; Greenberg & Koriat, 1991; Greenberg, Koriat, & Shapiro, 1992; Hadley & Healy, 1991; Healy, Conboy, & Drewnowski, 1987; Koriat & Greenberg, 1991). Based on the observations that alternating case type (e.g., Drewnowski & Healy, 1977), insertion of asterisks and spaces in continuous text (e.g., Healy et al., 1987), the introduction of misspellings (e.g., Healy, 1980), or the reversal of reading direction (Hadley & Healy, 1991) at least reduced the MLE, a unitization model of reading has been brought forward by Healy and her colleagues (Drewnowski & Healy, 1980; Hadley & Healy, 1991; Healy & Drewnowski, 1983). The unitization model like the perceptual frequency hypothesis of item-specific word form stresses the familiarity of the visual

pattern: “Familiar orthographic patterns enjoy fast activation of corresponding unitized representations, so that for these patterns, access to whole word entry wins the race of letter mediated access“ (Koriat & Greenberg, 1991). However, syntactic structure also seems to play an important role for the MLE: In a systematic series of binational studies Koriat, Greenberg and their colleagues, for example, observed that nonwords in the syntactic slot of a function word also produce an MLE (Koriat & Greenberg, 1991). Moreover, letter detection accuracy of the very same word in text seems to depend on its syntactic role in the sentence (Greenberg & Koriat, 1991; Koriat, Greenberg, & Goldshmid, 1991; for a review, see Greenberg et al., 1992).

The MLE was used by Nißlein et al. (1999) to investigate how distortion of word form in German words affected the missing letter effect. In three experiments they found that the manipulation of case of the initial letter (Experiment 1) or of the subsequent letters (Experiment 2) or manipulations of size (Experiment 3) all reduced the MLE. A particular reduction, namely an absence of the MLE was found, when the case of the initial letter was manipulated like in my experiments. Nißlein et al. (1999) concluded from the differences in reduction of the MLE that both visual word form and word class played a role for the MLE.

1.3. What is Learned with Regards to Visual Word Form?

Perceptual learning was found to occur at various levels of the visual system and the degree of specificity varies with the complexity of the training conditions (Ahissar & Hochstein, 1996, 1997). The perceptual frequency hypothesis of item-specific word form does not imply that perceptual learning is *exclusively* item-specific, but only that it is *also* item-specific (see discussion of the implications of Experiment 4 below). An exclusively item-specific perceptual learning hypothesis is, in my view, difficult to state, because nonwords never seen before can also be affected by word form distortions (e.g., Besner & McCann, 1987; Kinoshita, 1987; Mayall & Humphreys, 1996a). However, an exclusively general abstract account of visual word form is also hard to maintain because I obtained differential effects for the same word form manipulations in different word groups, normally presented in different perceptual form. Thus, it seems that it is indeed the degree of specificity which seems to vary with the complexity of the presentation conditions within one language. For example, in English case-specific letter encoding and abstract letter encoding are hard to disentangle, because almost all common words are presented very specifically in lower-case form, except for sentence beginnings and titles. Thus, claims about lexical access being based on abstract letter identities (Mayall et al., 1997; Paap et al., 1994) may in fact be very much based on specific language properties. When Mayall et al. (1997) postulate that „lexical access is

based on units coded at multiple levels [..] (that are same size and same case)" (p. 1285) on the basis of an examination of the English language in which all words are usually presented in same size, same case (namely small case), then obviously the question arises whether this is only true for English as a special case or whether it is a universal translingual phenomenon. The data of Study 1 clearly showed that same case, same size presentation did not lead to superior results for German nouns (in contrast to German non-nouns). Hence, the answer to the question what perceptual units are learned and used in visual word recognition may have both language-specific and translingual components. In other areas of visual word recognition, namely phonological encoding in dyslexics and normal children , significant differences in what is learned and what not already have been observed (Landerl, Wimmer, & Frith, 1997; Wimmer & Goswami, 1994). My results indicate that for visual word form perception similar principles hold.

The important question of what units are learned should perhaps be split up into two issues:

1. What units are learned?
2. Are these units abstract or item-specific or both?

Sometimes, the two issues cannot be distinguished, namely when units are learned that code whole word forms. For example, Haber et al. (1983) claimed that the envelope shape of a word is an important source of information in reading. Clearly, in this case the two questions fall together: The unit 'whole word envelope shape' is by its very nature item-specific for that particular word. Therefore, when Paap et al. (1984) convincingly demonstrated that envelope shape does not seem to play much of a role when the confounded changes in letter shape are systematically controlled, they addressed both issues simultaneously: Envelope shape is not learned and, hence, for that unit no item-specific word form effect could be observed. However, with regards to other units, such as letters, the issues can be separated. The answer to question 1 is clear: Letter identities can be learned and they *are* learned according to any current model of visual word recognition. But are *only* abstract letter identities learned (Mayall et al., 1997; Paap et al., 1984) or are *also* item-specific, case-specific letter identities learned?

The results of Study 1 indicate that an abstract letter identity approach is not sufficient. However, if the assumption that for specific items case-specific letter identities are learned is taken into account, then my data can be explained in a straightforward manner. I would hypothesize that for letter units both forms are learned, an abstract letter identity, as well as a case-specific, item-specific identity. This would imply that for cognitive psychologists 'mcclelland' is more difficult to recognize than 'McClelland', but due to using abstract letter identities 'mcclelland' can still be recognized. However, this effect is suggested to be item-

specific: ‘RuMelhart’ may not be recognized better than ‘rumelhart’. The use of item-specific, case-specific letter identities would also be consistent with findings showing that acronyms like FBI are sometimes better recognized in upper-case (see Besner, Davelaar, Alcott, & Perry, 1984; Experiment 1 for RVF; Henderson & Chard, 1976).

However, letter identity and envelope shape are not the only units or features that have been suggested to influence visual word form processing (see introduction and Mayall et al., 1997; Paap et al., 1984). According to a multiple levels approach of perceptual learning, transletter features of familiar letter groups may be processed in both ways, item-specific and abstract. Whether visual features are processed item-specifically or abstractly should - according to the perceptual learning research - depend on the specificity of the training conditions (Ahissar & Hochstein, 1997). The less specific the perceptual frequency of a particular feature, the less item-specific the processing of this feature may be. For example, the transletter group “Mc“ with ‘M’ in upper-case and ‘c’ in lower-case is not an item-specific feature of ‘McClelland’, as this particular case combination is perceived consistently over many surnames. However, as ‘Mc’ is most frequently perceived in that very case combination, the perceptual frequency hypothesis suggests that any name starting with the letter combination ‘Mc’ is best recognized when ‘Mc’ is presented in that particular case combination. In sum, I suggest that at any level of visual word form processing, a unit is generally best recognized in the visual form it is most often encountered in, but that the reference frame of the phrase “most often perceived“ is both, item-specific and item-overlapping, i.e., “abstract“, when the reference frame overlaps all items.

In other areas of visual word recognition both, frequency and specificity of visual units, are already known to influence performance. As concerns orthographic and phonographic sublexical units like bigrams, rimes or subsyllabic components, for example, a high number and frequency of those units are known to facilitate recognition (Massaro & Cohen, 1994; Massaro, Jastrzembski, & Lucas, 1981; Nuerk, Rey, et al., 2000; Ziegler & Perry, 1998), while, in contrast, little specificity (i.e., inconsistency) hurts the perception of print (e.g., Jared et al., 1990; Stone, et al., 1997; Treiman et al., 1995; Ziegler et al., 1997). I suggest that the same principle holds for the visual form of units involved in visual word recognition.

So far, I have argued that the units in visual word form perception can be both abstract and item-specific, but I left open the question, whether there are perceptual units that are exclusively (and not also) item-specific. In particular, are there whole word units or word-specific visual patterns (Drewnowski & Healy, 1980; Hadley & Healy, 1991) or not (Besner, 1989; Mayall et al., 1997; Paap et al., 1984)?

Currently, I believe this issue is still open, because AFL that was thought to deny the role of item-specific word form processing is theoretically questionable and empirically not supported by my results.

1.4. On the Additivity Logic in the General Linear Model as Evidence against Item-Specific Processing

"The assumption that stage-durations are additive has often been incorporated with the idea that they are independent. [...] It is quite conceivable, however, that in some situations stage durations might be additive but not independent." (Sternberg, 1969, p. 302).

This issue has more general implications than the issue of item-specific perceptual learning in word recognition. I argue that denials of perceptual and cognitive processes based on the premise that the general linear model is true and on the conclusion that processes are additive and sequential, if a null interaction in an ANOVA is obtained, may be misleading. Null interactions in an ANOVA can have, at least, two possible causes: First, there may be no interaction between two or more additive factors. Second, the general linear model and AFL may not be appropriate for analyzing that particular problem, while other accounts, such as interactive activation frameworks, are more appropriate.

Although Sternberg (1969) questioned the independence of additive factors even within constrained (sequential) stage-based models, the usual arguments against item-specific visual word form processing just rely on that assumption: Additivity within the much less constrained (e.g., not sequential) general linear model is taken to prove independence of the involved factors. Typically, the logic described in detail in the introduction goes as follows: Visual distortion (e.g., case mixing) and frequency / lexicality (words vs. nonwords, or words of different frequencies) are manipulated in a two-factorial design. Then, mostly approximately, additive effects of the visual distortion and frequency are reported, but no interaction (e.g., Besner & McCann, 1987; Frederiksen, 1978; Mayall & Humphreys, 1996a). Based on that null interaction, it is concluded that item-specific visual form does not play any role. The argument is that if the item-specific word form would play any role, then words of higher frequency should be stronger affected.

I argued in the general introduction that this line of reasoning is problematic because it does not take into account lexical top-down processes (stronger for higher lexical frequency) that work into opposite directions than perceptual frequency processes. As a consequence, additive effects can be produced in interactive models, too. Hence, even in the presence of item-specific word form processing, one might find additive effects. The results of Experiment 3 support this claim: In the presence of item-specific word form processing (i.e., an interaction of perceptual frequency of word form with the presented initial case

manipulation), I found additive effects of frequency and stimulus manipulation for both, nouns and non-nouns. Thus, if I used the additivity argument, I should conclude that there are no item-specific effects, while the diverging results for nouns and non-nouns clearly demonstrate such item-specific effects.

Thus, the present study empirically supports the argument that applying additivity logic to null interactions in an ANOVA can be misleading and thereby questions the validity of this logic not only for the elimination of the idea of item-specific visual word form processing, but more generally. The ANOVA relies on a statistical and theoretical model, the general linear model which assumes that data can be adequately described as a linear additive combination of independent factors and their interactions. However, other models as, for example, cascade or interactive models (Jacobs, et al., 1998; McClelland, 1979; Plaut et al., 1996; Rumelhart & McClelland, 1982; Zorzi, et al., 1998) have given up a linear additive independent factor logic, and can, nevertheless also produce additive effects (e.g., Rumelhart & McClelland, 1982, Exp. 7).

Therefore, at least in fields, in which interactive models compete with sequential stage models, null interactions in an ANOVA interpreted within the framework of sequential stage models are not conclusive. Although the general linear model is -as the basis of the ANOVA- probably the theoretical and statistical model most often assumed to be true, this is not necessarily the case. This objection does by no means imply that the ANOVA is an inappropriate inference-statistical model that should no longer be used. I used ANOVAs throughout this study and, for inference statistics in field of visual word reocgnition, it may still be one of the best horses in town. However, I must keep in mind that its theoretical foundation, the general linear model, is still a model and any ANOVA effect or null effect can only be interpreted along the assumption that the general linear model is true. Otherwise, ANOVA results can be misleading. This study represents an example for a field in which this may have been the case.

1.5. The Perceptual Frequency Hypothesis of Location

Study 2 provided further support for the perceptual frequency hypothesis because the asymmetry of the viewing position curve shifted as a function of word length. Longer words produced more asymmetrical viewing position curves than short words. This shift corresponds to the perceptual frequency of landing positions for longer and shorter words (Nazir, 2000; Nazir et al., 1998; Vitu et al., 1990). That landing positions and the asymmetry of reading performance are so closely related for different word lengths can be seen as evidence for the perceptual learning hypothesis because this is inconsistent with other accounts of the asymmetry of the viewing position curve. Alternatively, one could

have argued that the asymmetric shift to the left could reflect a general attribute of reading in languages with left-to-right direction because I only observed a correlative relation between fixation distribution and performance. It is not clear whether this correlation is accidental and modulated by other determinants of word recognition, as for example, sequential grapheme-phoneme conversion which could operate from left to right (Rastle & Coltheart, 1999).

However, the manipulation of word length in Study 2 reflects an experimental test of the perceptual frequency hypothesis of location. On the basis of different saccade landing distributions for different word lengths, the perceptual frequency hypothesis predicts testable and falsifiable which viewing position curve shapes should be expected: More symmetrical viewing position curves for short lengths and less symmetrical curves for longer words. Study 2 showed that this was indeed the case. Thus, Study 2 represents a successful experimental test of the perceptual frequency hypothesis of location. Taken together with Study 1, these data strongly suggest an important role of elementary visual-perceptual processes in visual word recognition.

1.6. On the Relation between the Perceptual Frequency Hypotheses of Location and Word Form

If perceptual frequency of location influences visual word recognition at specific locations, then it should do so particularly at the most frequently fixated locations. For most words, these locations are on the central fixation positions. If one further assumes an interactive account of visual word recognition, then the current results would consequently lead to the following prediction: Perceptual frequency is most likely to mask other effects of visual word recognition at those positions and for those word and letter forms that are perceived most frequently. Indeed, there was some indication for such an interaction in my data. In Experiment 2, ALC and IUC forms of non-nouns were particularly better recognized on the initial positions. In German, the individual letters and their features are commonly perceived in lower-case at the end of the word, while at the beginning of the word capitalization is possible and occurs inconsistently. Thus, one might look at an interaction of „where“ and „how“ features have most often been seen in the past: What is most often seen are features of lower-case letters to the right of fixation. In contrast, features of upper-case letters are less often seen to the right of fixation. Finally, to the left of fixation both features of lower-case and upper-case letters may be perceived. According to the perceptual learning hypothesis, one would expect superior performance at the location and in the form features are most often and most consistently seen, i.e., when features of lower-case letters are seen to the right of fixation. In Experiment 2, this was the case in the ALC and IUC condition.

1.7. Conclusions

Altogether this thesis presents evidence for the important role of perceptual frequency in visual word recognition from three different effects: The visual word form effect found task-overlapping in Study 1, the modulation of the optimal viewing position by word length in Study 2, and the interaction of the perceptual frequency hypothesis of location and word form in Study 1. The perceptual frequency hypothesis accounts for the word form effects by assuming that a word is best recognized in the particular form it is most often perceived. The perceptual frequency hypothesis accounts for the word form effects by assuming that a word is best recognized in the particular form it is most often perceived. As this effect is item-specific, general, abstract accounts of the effect are not sufficient. Experiment 3, in particular, showed that null interactions of word form effects with lexical frequency are not conclusive when they are interpreted in the general linear model using AFL as indicating that item-specific word form does not play any role. In the same Experiment 3, with appropriate manipulations, item-specific word form effects were found. Thus, rejections of item-specific word form effects based on null interactions in the ANOVA may be premature, because they rely on an AFL within the framework of the general linear model that may not be appropriate to investigate complex dynamic systems (Bosman & Van Orden, 1997; Van Orden & Goldinger, 1994). With regard to connectionist models of word recognition, these results suggest that it is not sufficient to implement visual word form manipulations just as detrimental noise, because I found a word form manipulation that is not detrimental to all stimuli but that can produce superior performance in one word group and inferior performance in another.

Second, the perceptual frequency hypothesis of location predicts a systematic shift of optimal viewing position with word length, as the saccade landing distribution is altered by word length. Experiment 4 showed that the shift of the VPE followed the shift of fixation distribution. Any account of the VPE independent of word length will have difficulties to integrate these data. Finally, the interaction of the above two perceptual frequency effects was indicated in Study 1. Word form effects were stronger at more frequently fixated position. However, as this effect was found in Experiment 2 but not Experiment 1 more research is needed to reliably confirm this interaction. In sum, this thesis presents strong evidence for an important role of perceptual frequency of location and word form in visual word recognition. More generally, these data indicate that the abstract case-unspecific letter code of all current models of visual word recognition is, by the very nature of its input code, unable to simulate item-specific word form effects. Thus, these findings represent a challenge for current models of visual word recognition.

2. Visual Alteration of Lexical Processes as Indexed by Frequency

2.1. Introduction

Lexical frequency effects and its interactions with other variables can be altered by visual presentation conditions. In this general discussion, I will argue that this alteration should be considered because under near-optimal presentation conditions null interactions with other variables are particularly likely. However, such null interactions may not generalize to other presentation conditions, but may be specific to near-optimal conditions, in particular to fixating on central viewing position with unlimited presentation time.

2.2. On the Generalization of Null Interactions of Frequency with Other Variables

The controversy about whether frequency interacts with other variables, such as visual quality, is usually empirically investigated in factorial experiments with central (near optimal) viewing position, under good laboratory presentation conditions, and in skilled participants using an ANOVA based on central tendencies in the general linear model. Null interactions of frequency with other variables may be due to all or any of these favorable conditions. My experiments together with older studies present evidence that these conditions may mask frequency effects and their interactions (see general introduction for a detailed review). As a consequence, word recognition models which are partially based on such null interactions (e.g., Borowsky & Besner, 1993) may present an overgeneralization. I will elaborate on this argument by integrating my findings and those of previous studies in more detail next.

First, my results in Study 2 show that frequency effects were smaller for a central (near optimal) viewing position than for edge viewing positions in the PIT. These results converge with earlier findings of O'Regan and Jacobs (1992) in the LDT and the NT as well as with the results of some eye movement studies (McConkie et al., 1989; Pynte, 1996). Moreover, it is also in line with results that orthographic-lexical effects tend to be strongest farther away from fixation (Grainger et al., 1992, and Study 3 in which, however, the frequency by viewing position or condition interaction did not reach significance in the ANOVA). For frequency as an index of lexical (or post-lexical, cf. Paap & Johansen, 1994) processes, the hypothesis seems to hold that its interactions with other variables are most likely to be missed on central fixation. However, this is the condition used in standard experiments which are generally interpreted with respect to the specific presentation condition. Let me consider the interaction with the word length effect to illustrate this issue. In my study, the word length by frequency interaction is particularly present on edge viewing positions compared to the central viewing position. Long, low-frequency words produce particularly inferior performance on poor viewing positions. This finding questions interpretations of null

interactions between frequency and word length in central viewing position as being general for word recognition.

Similarly, frequency itself may be affected by visual quality of laboratory stimulus presentations. In a word fragmentation task (cf. Snodgrass & Mintzer, 1993; Snodgrass & Poster, 1992) I found frequency to be more of a factor for recognition level in words with low letter legibility and high confusability.[15] Concerning the question as to whether an interaction between frequency and visual quality exists or not, I doubt that one can find a general answer, but rather only one dependent on the circumstances. In particular, the answer seems to depend on the specific laboratory degradation or distortion procedure, stimulus attributes and manipulation methods (see Allen et al., 1995; Kinoshita, 1987; Norris, 1984; Wilding, 1988 for similar effects, but Becker & Killion, 1977; Besner, 1989; Besner & McCann, 1987; Borowsky & Besner, 1993 for diverging effects). In this context, recall again the findings of Study 1. While no interaction of perceptual and lexical frequency was obtained in Experiment 3, the item-specific visual word form nevertheless played an important role in facilitating or inhibiting LDT reaction times. Here, item-specific visual word form as a procedure of visual manipulation led to perceptual frequency effects that could not be observed in interaction of this visual variable with frequency on central fixation. Another example of the dependency on manipulation comes from the comparison between Studies 2 and 3. While in Study 2 frequency effects increased with inferior viewing position in every possible analysis, a similar effect was not significant in Study 3. A likely reason for this divergence is that the visual manipulation in Study 3 was not strong enough and that the variance of overall performance was too small to detect any frequency by viewing position interaction there. Again, if I had only interpreted the null effect in Study 3, I would have come to premature conclusions. The interaction between frequency and other variables may be detected only if the visual manipulations are carefully chosen and their effect is strong enough to produce a large variance of overall performance,. Therefore, null effects based on arbitrary, albeit possibly weak, visual manipulations do not help much to clear up the nature of the interaction of visual and lexical processes.

A similar argument holds when, instead of stimulus or presentation attributes, participants' attributes are non-optimal. As regards normal and beginning readers, the frequency effect and the word length

[15] Note that differences in neighborhood confusability are not confounded with the interaction between frequency and letter confusability in Experiment 5. The neighborhood confusability was generally on a middle level and words with generally highly confusable letters not surprisingly tended to have a higher N-conf index, too. However, even this small difference in N-conf between low and high LCI words did not differ between frequency groups. Due to this additivity an interaction of letter confusability and frequency is not due to simple stimulus differences in neighborhood confusability. The confusability effect itself in Experiment 5 was replicated in Experiment 7 for hermits, so that general letter confusability effects are not fully explained by neighborhood confusability.

effect and its interactions seem generally to play more of a role when the reading capabilities are non-optimal or severely damaged as in alectics (Aghababian & Nazir, 2000; Behrmann et al., 1998; Butler & Hains, 1979; Montant et al., 1998; Perfetti et al., 1979). Optimal reading skills may mask these effects in a similar way as, for example, on the stimulus side high-frequency tends to mask orthographic-phonological recoding effects which are mostly observed for low-frequency words only (see, e.g., Backman, Bruck, Hebert, & Seidenberg, 1984; Jared & Seidenberg, 1991; Nuerk, Rey, et al., 2000; Rey, Jacobs, Schmidt-Weigand, & Ziegler 1998; Seidenberg, Waters, Barnes, & Tanenhaus, 1984; Ziegler & Perry, 1998; but Jared, 1997, for an exception).

Finally, null effects of frequency and null interactions of frequency with other variables based on measures of central tendencies in the ANOVA may not be conclusive because such effects may still be detected with more fine-grain analysis. Recent distributional studies (Balota & Spieler, 1999; Plourde & Besner, 1997; Schmidt-Weigand et al., 1998) show that frequency does not only exert its effect on central tendencies which are statistically tested in the ANOVA. When using the ex-gaussian deconvolution, all of these studies show that linguistic variables, such as frequency, do in some tasks not only influence the central tendency of the distribution, but also the shape of the distribution. In particular, frequency tends to affect the exponential τ component of the ex-gaussian deconvolution, and thus the skewness of the distribution in the studies of Balota and Spieler (1999) and Plourde and Besner (1997; see introduction for details). I conclude with the authors (see also Robert & Sternberg, 1993; Sternberg, 1969) that investigating measures of central tendencies in an ANOVA may not be sufficient to test sequential stage models against interactive models.

Therefore, failure to find frequency effects and their interactions with other variables under near-optimal presentation conditions and in properly skilled participants in an ANOVA over measures of central tendencies based on the general linear model may produce a misleading beta-error for any of these reasons. In particular, any such null effect obtained under near-optimal presentation conditions (such as central viewing position) should be tested for its general reliability under non-optimal presentation conditions (such as viewing position at the beginning or the edge of a word).

2.3. Consequences for Models of Visual Word Recognition

The findings in this thesis have some clear implications for models of visual word recognition. First, as interactions of lexical with visual presentation conditions were observed here as well as in other studies (Beauvillain, et al. 1996; Brysbaert, et al., 1996; Hyönä, 1995; O'Regan, et al. 1984; Underwood, et al.,

1990), models need to take into account visual variables to account for those effects (see, e.g., Jacobs et al., 1998; but in contrast, Plaut et al., 1996). The argument that such interactions are beyond the scope of a respective model does not hold if the model should account for lexical processes. A null interaction of frequency with another variable (such as word length in my study or visual quality in other studies) under near-optimal presentation conditions may be particularly critical. Following the AFL (Sternberg, 1969), models may generally assume independent and non-interacting stages of processing because of null interactions that may simply be specific to laboratory conditions.

This interaction between frequency and viewing position (and word length) also sets clear restrictions on any further development of the model of Nazir et al. (1998). An independent additive lexical height parameter, as it was planned in earlier stages of the model (Nazir, 1991), cannot by its very nature account for the interaction of frequency with viewing position. Any lexical frequency parameter newly introduced must interact with viewing position in a specific way. As long as word recognition is not virtually impossible, frequency must exert its effect particularly on viewing positions on which performance is low, and particularly for long words. A suggestion for the specific mathematical incorporation of “as long as word recognition is not virtually impossible” will be given in the model refinement chapter.

3. On the Alteration of the Word Length Effect

In the general introduction I reviewed the literature suggesting that word length effects are most likely to occur when stimulus attributes, presentation conditions and participant skills are non-optimal. In Study 2, in which I studied the word length effect, stimulus attributes (frequency) and presentation conditions (viewing position) were varied to test this hypothesis.

For both manipulations I found stronger word length effects under non-optimal conditions. As regards the word length effect, I observed the same pattern as for the frequency effect – it tended to be weakest on optimal or near-optimal fixation positions and stronger for poor fixation positions, i.e., the beginning or the end of a word rather than the middle. Performance decreased linearly with increasing word length on edge viewing positions but not on central viewing position. Even if no linear trend was assumed in the statistical analysis and a simple comparison of word length is made in t-tests, longer words were slower than shorter words on edge positions but not on central viewing position. So, clearly, the word length effect is stronger or only obtained under unfavorable viewing conditions. In my opinion, this strongly questions general conclusions such as those of Weekes (1997, p 439): *“Number of letters continued to affect nonword naming latency, but not low-frequency word naming latency, after the effects of*

orthographic neighborhood size, number of friends, and average grapheme frequency had been accounted for." This statement may be specific to Weekes' (and the standard research) procedure that *"stimuli were presented centrally (p. 443)"* rather than on different viewing positions. Null effects of word length may be a laboratory artifact of central viewing position. Thus, the general interpretation of such a central viewing position result within the framework of word recognition models as the DRC (Coltheart et al., 1993; Rastle & Coltheart, 1999) may be premature. The same arguments hold for task differences or task difficulty findings (e.g., Hudson & Bergman, 1985).

The second alteration I investigated was the alteration of the interaction of the word length and frequency effects. I have already discussed the nature of this interaction of frequency and word length in the last chapter with respect to the alteration of frequency. Generally, one observes overadditivity for the two effects. For the word length effect, this implies that it is greater in poor, i.e., low-frequency conditions (while vice versa the frequency effect is greater for poor, i.e., long words). As noted above, it is particularly dangerous to interpret a null interaction of word length with frequency under optimal viewing conditions, as I found such an interaction on edge viewing positions, but not on central viewing position. Again, this has significant consequences for modeling this hallmark effect of visual word recognition. As the nature of this effect and its interactions seem to vary strongly with viewing position, results of one particular viewing position can again not be generalized on the natural reading process.

For this thesis, I have generalized the model of Nazir et al. (1998) such, that it is capable of fitting different word lengths (with an unequal number of letters and viewing positions). The fitting results imply certain constraints for the refinement of the model. While for three- and five-letter words, the fits are good, the fits for seven-letter words strongly deviate from the empirical curves. The fitted curves are too flat, especially on the edge viewing positions for long words. The model underestimates the drop-off in overall performance for long words. One reason for this underestimation may be that a linear drop-off rate of letter legibility probability may no longer be appropriate for long words as negative probability values would become possible. To prevent such negative probability values the drop-off rates to the left and right stay fairly low, so that even the lowest legibility of a letter at the poorest position is positive. A second reason might be that a frequency parameter has not been introduced yet. Evidence from the literature and my own findings suggest that frequency exerts a particularly strong effect on long words. The behavior of the model without the parameter might be an interesting example of how the lack of this parameter affects the performance of the model in a non-linear, non-additive way. Without the frequency parameter, the model tries to catch the mean height of the curves by varying its drop-off rate. If one

argued simply additively and linearly (with AFL using the general linear model), the frequency parameter should improve performance most on inferior viewing positions. If this was the case, a model without a frequency parameter should consequently underestimate performance particularly on poor viewing positions. However, the opposite is the case. The reason for this apparent contradiction is that the drop-off rate is the sole parameter for height and shape of the curve (besides asymmetry). The empirical curves obtained for long words can simply not be produced by the two-parametrical model as it is now. A frequency parameter that allows a broader range of curve forms might allow better fits for longer words, too, and the greater variation of shapes induced by this parameter seems to be most necessary for long, seven-letter words for which frequency effects were greatest.

Finally, the drop-off parameters for the different word lengths were not invariant: For shorter words they were much larger. The reason is that the probability of recognizing the letter on fixation is set to 1. Thus, in three-letter words there are only two letters left which may not well recognized. When the old model fitted inferior performance (computed as a product of the simple letter legibilities), hence, these two remaining letter legibilities had to be assumed to be very low which could only be achieved by assuming very high drop-off rates. The consequences for the refined model are that the legibility likelihood for the letter on fixation can no longer be set to 1 if words of different length are fitted with one invariant drop-off parameter.

Taken together, the fits of the model underline the central message of this general discussion: Word recognition effects cannot generally be studied or modeled with the principle of isolated variation, because they interact in a complex nonlinear way with each other. Empirically, the word length effect itself and its interactions may be underestimated under optimal presentation conditions (viewing position) and stimulus attributes (frequency). The modeling of a viewing-position-dependent word length effect seems to be particularly difficult in the absence of a lexical (frequency) parameter. Hence, one must be careful in generalizing null effects of word length obtained under near-optimal stimulus or presentation conditions.

4. Neighborhood as an Index of Similarity in Visual-Orthographic-Lexical Space

4.1. Introduction

The visual-orthographic-lexical space in the written word universe does not consist of equivalent visual hyperspaces of orthographic-lexical similarity. Similarly to the frequency effect and the word length effect, the results of this thesis do also indicate that the neighborhood size effect is strongly modulated by visual

determinants. This discussion will focus on three issues. First, the neglect of visual properties in examining neighbors as index for orthographic-lexical processing will be critically discussed. Second, as a consequence of the first issue, the examination of so called Body Neighbors without respect to their positional information distribution will be analyzed. Third, the implications for modeling and, in particular, the refinement of the Nazir et al. (1998) model will be examined.

4.2. The Number of Neighbors as Index of Similarity in Visual-Orthographic-Lexical Space

Study 3 demonstrated that at least in PITs N is better understood as indexing visual-orthographic-lexical similarity rather than pure orthographic-lexical similarity. Thus, the findings of Grainger et al. (1992) obtained for two fixation positions with fixation only on or off higher frequency neighbors can be generalized to all fixation positions and to all kinds of neighbors. Neighbors on letter positions that were far away from fixation and thus not well legible tended to be particularly detrimental to performance. In a similar way, disambiguating letters of two neighbors seemed to be particularly inhibitory when they were visually highly confusable. Reviews of the neighborhood effect (Andrews, 1997) that do not discuss neighborhood as a visual-orthographic-lexical effect may underestimate other modulations of the neighborhood effect that are confounded with visual determinants.

Moreover, Experiment 7 pointed towards a possible confound between neighborhood effects and the dependent variable measured. Accuracy seemed to be more sensitive for inhibition effects due to neighborhood competition, while recognition speed is more sensitive to facilitatory effects possibly due to more frequent orthographic-phonological correspondences in words with many neighbors. Consequently, results obtained with a given dependent variable (such as RT in the NT or the LDT) cannot be generalized to all dependent variables and, in particular, not to the reading process in general. Models that are only able to account for one dependent variable may thus not be able to simulate or fit the whole scope of neighborhood effects.

In sum, the most favorable conditions for facilitatory neighborhood effects seem to be very high letter legibility (or low confusability), reaction speed as a dependent performance variable and, possibly, a task in which sublexical orthographic-phonological correspondences as facilitatory components may play a major role (possibly the NT). Thus, the standard laboratory task in which a single word is presented on central fixation position under good illumination, seems to enhance the facilitatory attributes of neighbors. This conclusion is also confirmed by recent experiments of Pollatsek, Perea, and Binder (1999). They

found that N had a facilitatory effect on word recognition in a laboratory LDT, while in normal[16] text reading with the same words N produced an inhibitory effect that was due to longer gaze duration on words with larger N. Together with findings from eye movement studies that had indicated that information ambiguity within a word interacted with gaze duration (e.g., Beauvillain et al., 1996; O'Regan et al., 1984), these results strengthen the hypothesis that N effects in single word recognition tasks with central fixation may not be conclusive. Near-optimal presentation conditions in laboratory tasks may lead to fundamentally different effects than observed in normal reading. Hence, conclusions based on such presentation conditions, such as there being no lexical lateral inhibition between neighbors (cf. Andrews, 1997), are not easily generalizeable, but may be restricted to those presentation conditions.

4.3. An Example for a Possible Confound between Visual Determinants and Orthographic-Phonological-Lexical Variables

An example for a possible confound between visual determinants and other effects may be the investigation of so called Body Neighbors (BodyN effect). This example may show that holding visual attributes constant at a certain presentation condition does not necessarily allow us to generalize the results from this presentation condition to any other. BodyN is defined as the number of words that share the same body or rime with the target word (e.g., Colombo, 1986; Ziegler & Perry, 1998). Targets with a high BodyN tend to be recognized faster than control words. The standard interpretation of this result is that bodies are important sublexical orthographic-phonological units and that word recognition performance improves if its constituent sublexical units are frequent (e.g., Ziegler & Perry, 1998). However, although Body Neighbors do not need to have the same word length as the target word, this often seems to be the case, and hence, body neighbors are often also orthographic neighbors. Therefore, words with a high BodyN, but a similar number of neighbors, tend to have their disambiguating information at the beginning of the word rather than at the end. Control words used in these studies consequently, tend to have their disambiguating information at the end rather than at the beginning because their neighbors must not be BodyN. In sum, words with a high BodyN tend to have their neighbors on the initial positions and their control words on the final positions.

This positional distribution relates to the findings in Study 3. Viewing position interacted with neighborhood position in a characteristic way. The farther neighbors were away from fixation, the greater

[16] "Normal" text reading is still a laboratory task using biteboards, thus, certainly still being not an equivalent of free reading. However, the point is that when words are no longer presented centrally under near-optimal presentation condition in a task that is only used in word recognition experiments, the results already differ from "normal" text reading even with the use of biteboards.

was their inhibitory effect. However, the interaction between viewing position and fixation was asymmetrical - the viewing position curve became symmetrical for words with disambiguating information at the end, but heavily asymmetrical for words with disambiguating information at the beginning. Particularly important for studies with central viewing position is that words with their neighbors on the initial positions also produced superior performance on central viewing position. Thus, results obtained with central viewing position may not be generalized to all viewing positions, but may particularly emerge for initial and central viewing position as a consequence of this interaction. When participants fixate at the end of words, such effects may disappear. For BodyN, this may imply that a superiority of words with many BodyN compared to control words may only emerge on central viewing position, but not on other viewing positions. As in natural reading, where fixation position is distributed over all spatial positions within a word (Nazir et al., 1998; Rayner, 1979; Vitu et al., 1990), any result obtained with central fixation position must be carefully examined before it is generalized for the normal reading process. In particular, it should be examined whether superior performance for words with a high BodyN are due to their sublexical orthographic-phonological properties or whether it is only due to the specific interaction of their particular neighborhood distribution with (central) viewing position compared to control words. I suggest that this confound should be disentangled in future studies by manipulating fixation position and examining the interaction of this hypothesized orthographic-phonological property with visual properties.

4.4. Implications for Models Simulating or Fitting Orthographic Effects

The results of my thesis also indicate that neighborhood effects can hardly be modelled or fitted in general without reference to the visual alteration of these effects. Any model without visual features will in my opinion fall short of explaining these visual-orthographic interactions at least in PITs. Therefore, interactive activation models, which have a visual feature level (Grainger & Jacobs, 1996; Jacobs et al., 1998; McClelland & Rumelhart, 1981) are architecturally better suited to simulate such effects than models which lack such a level (e.g., Plaut et al., 1996). Furthermore, as mentioned above, models that are able to simulate only one dependent variable are less suited to model neighborhood effects than models that can simulate both RT and accuracy data because neighborhood size effects may change as a function of the dependent variable. However, even this interaction strongly depended on the visual properties of the neighbors in my study. Therefore, the visual conditions under which a model operates must be specified. If it is only able to operate under one specific visual presentation condition, the processes the model uses to simulate neighborhood effects are questioned in their generality.

With regard to the refinement of the model of Nazir et al. (1998), the asymmetry parameter could be influenced by the distribution of lexical information ambiguity implying that the asymmetry parameter cannot be indexed as a purely visual parameter. The purely visual notion of the asymmetry parameter as an asymmetry ratio in the drop-off rate grounding on an empirically observed asymmetry in letter legibility (Nazir et al., 1992) is no longer sufficient if neighborhood effects are to be included in the scope of a refined model. Vice versa, it is not sufficient to take N into account as an additive inhibitory variable because N interacted with visual parameters like the asymmetry and drop-off rate. A refined model must therefore take into account that local information ambiguity of a certain word affects performance differently on different viewing positions. The viewing position curve of a word is hence not only mediated by individual properties of a given word, but also by its relation to other words in the specific orthographic-lexical hyperspaces for different visual conditions (i.e., different viewing positions).

5. A Refined Nonlinear Arithmetical Model of Word Recognition

5.1. Problems and Restrictions in the Scope of the Current Model

Based on the constraints imposed by the results of Studies 1-3, a refined arithmetical model shall now be developed. In particular, the goal is to make the step with regard to the scope of the model: The viewing position model of Nazir et al. (1998) which was developed to fit viewing position curves should now become a real word recognition model, in which words with different stimulus attributes on different viewing positions can be fitted with an invariant set of parameters. Recall, for example, that the model was only able to fit a frequency effect by altering visual parameters such as the drop-off rate. Other issues that must be considered in a refined model have emerged when the model fits were analyzed and discussed in the different studies. The following list gives an overview over the issues or shortcomings which should be considered for model refinement. Note that although some issues may refer to the same parameter in the model of Nazir et al., they nevertheless impose separate problems for a new model. For example, the asymmetry parameter may change as a function of word length (issue 2) and of neighborhood distribution (issue 3).

1. The model could produce negative word recognition values.
2. The asymmetry parameter (originally interpreted as a purely visual parameter) changed as a function of word length.
3. The asymmetry parameter (originally interpreted as a purely visual parameter) changed as a function of neighborhood distribution.

4. The visual drop-off parameter changed as a function of frequency.

5. The visual drop-off parameter changed as a function of word length.

6. For long words, the fitted curves were too flat.

7. The ratio of fitting five data points with two free parameters is already not a good ratio – adding further parameters only makes sense when the number of data points fitted is increased.

8. The probability of letter recognition on fixation is fixed to 1.

I will now elaborate on these points in greater detail.

Ad 1.) Consider the drop-off values in Experiment 4 (Table 8). In the current model, recognition probability decreases from 1 with increasing letter distance from fixation (with the gradient drop-off rate / letter distance). If the huge drop-off rate obtained for three-letter words were used for seven-letter words, I would obtain negative letter recognition values. That the model is capable of producing mathematically impossible word and letter recognition probabilities is clearly a shortcoming and needs to be resolved.

Ad 2.) That the asymmetry of the viewing position curve changed with word length is in line with the perceptual frequency hypothesis. Shorter words are more frequently fixated near or right from the middle of the word than longer words and should hence be better recognized in the middle of the word than longer words (see discussion of Experiment 4). However, the asymmetry parameter in the current model refers to an asymmetry in letter recognition when the letter is presented left or right of fixation (Nazir et al., 1992).This perceptual asymmetry of recognizing any given letter to the left or to the right should not change with word length, while the perceptual frequency of recognizing any given word at a particular position may change with word length.

Ad 3) The asymmetry parameter as a visual parameter should not change as a function of lexical information distribution within a word. Consider the results of Experiment 6. There asymmetry changed as a function of neighborhood distribution. Viewing position curves of words with their neighbors at the end were much more symmetrical than viewing position curves of words that had their neighbors at the beginning of a word. In a word recognition model that should be capable of fitting stimulus groups with different information distribution with invariant parameters, the visual asymmetry ratio should not differ.

Ad 4) Experiment 4 showed significantly and Experiment 6 showed tendencies that the shape and height of the curves may also change as a function of frequency. Because the visual drop-off rate is a

visual parameter, lexical or post-lexical attributes should not change their value if different frequency groups were fitted with invariant parameters.

Ad 5) Experiment 4 also showed that the drop-off parameter changed as a function of word length. As the drop-off parameter is thought to code the visual drop-off in letter legibility, this should not be the case. The new model should produce invariant drop-off rates for different word lengths. For short words, the drop-off parameter had to be especially high to obtain a low word recognition rate as the recognition probability at fixation was set to 1. Possibly, the simplifying assumption that letter recognition at fixation is set to 1 only holds for longer words (five letters and more), but no longer for shorter words.

Ad 6) The fitted curves for long words were too flat. The reason is that the drop-off rate is the only parameter that changes height and it changes shape at the very same moment. This implies that steep curves, such as those for long words in Experiment 4, could not be fitted at all with any parameter set in the current model. Curves with recognition values that are close to 1 or close to zero are flat, but curves with recognition values in between zero and 1 can fall from maximum to minimum performance only to a certain extent. Two things need to be done to solve this problem: A lexical parameter and a visual drop-off parameter that are fitted independently from each other should allow a greater variety of curve forms. Second, as the legibility likelihood of a letter decreases towards zero, the drop-off in legibility should be transformed in such a way that it cannot decrease under zero. Using this correction, a greater variety of drop-off rates becomes possible without obtaining negative recognition probabilities for a single letter and eventually for the whole word.

Ad 7) For practical purposes, at least within-participant designs are naturally restricted to a limited number of fixation positions because with an increasing number of fixation positions the number of data points for each condition at each viewing position decreases in an experiment with a fixed number of stimuli. In the experiments of this thesis, usually five or ten data points per position were obtained in each condition for each fixation position in one participant. If one wants to retain a sensible ratio between data points and additional parameters to be added in a refined model, one must increase the number of data points fitted with the very same parameter set. In my experiments, three or six conditions were presented on each fixation position. If these three or six data points per fixation could be fitted with an invariant set of parameters, the ratio between free parameters and data points would become much better. In particular, fitting different categories of words (i.e., words of different frequency, length, or neighborhood distribution) with the same invariant parameter set, would, in my opinion, mean a transformation from a visual model to a real word recognition model. Note that although the refined model for Experiment 4

could fit words for different word lengths, it could not do that within the same invariant parameter set. It treated the additional letters in longer words simply as additional visual information that needed to be picked up. However, as the visual drop-off rate for letter recognition likelihood varied greatly for different word lengths, this purely visual interpretation of word length effects falls short of the data.

Ad 8) That the probability of letter recognition on fixation is fixed to 1 is another simplifying yet previously successful assumption of the model. However, while this assumption works for longer words (five and seven letters), it did not work for shorter words (see problem 5). If legibility of a three-letter word is low, the model has to fit high drop-off rates to account for a low legibility of the whole word because the whole decrease in general performance must be obtained by the decrease in recognition accuracy of the remaining two letters off fixation. Only two out of three letters can be considered because the third letter on fixation is assumed to be recognized with probability 1. Therefore, this assumption did not work for three-letter words. A possible solution is that this parameter is allowed to vary freely between stimulus groups. However, again this only becomes possible when the number of data points fitted by the models with one invariant parameter set increases.

5.2. Goals for the Incorporation of the New Parameters

The foremost goal will be to keep the interpretation of parameters clear and simple, since models should be as simple as possible (cf. Jacobs & Grainger, 1994). In the old version of the model parameters that were thought to be visual parameters varied with lexical attributes of the presented words, e.g., the drop-off rate with frequency and the asymmetry ratio with neighborhood distribution. Thus, frequency effects could not well be parametrically interpreted and examined because the same parameter coded visual and lexical attributes. Hence, the goal for this model refinement is that the visual parameters should no longer vary with lexical attributes. Instead, different stimulus groups presented under identical visual conditions to the same participants should be fitted with the same parameter values. Furthermore, parameter values should be invariant with respect to manipulations that should not influence them. The lexical parameter, for example, should not differ as a function of word length, but only as a function of lexical variables such as frequency. This does not mean that the influence of frequency on word recognition performance is not allowed to differ in a systematic way. Rather, it means that the index of lexical processing in an individual should be invariant. Any systematic interaction with other variables should be adequately described by and be a result of the incorporation of the new parameters in the model. Specifically, the following invariance problems should be solved by the new model:

1. The asymmetry parameter should be invariant for different word lengths and neighborhood distributions.
2. The drop-off rate should be invariant for different word lengths and frequency groups.
3. The lexical parameter should be invariant for different word lengths and different viewing positions. Recall that in Experiment 4, frequency had greater effects on inferior viewing positions.
4. The parameter coding letter recognition probability on fixation should be invariant for different word lengths, different fixation positions, and different frequency groups.

Only when all these invariance goals have been obtained, can one successfully fit different data points from one experiment with one invariant parameter set. In the next section, a new model that may overcome the problems mentioned above is developed.

5.3. Towards a Nonlinear Arithmetical Model of Word Recognition (NAMWR)

The goal is to develop a revised model that accounts for the VPE, frequency effects, neighborhood distribution effects, word length effects, and their interactions with one another simultaneously while remaining clearly testable and descriptively adequate. The goal is to present an alternative to criticized general linear model in the specific domain of word recognition. The NAMWR may give up the assumption of linearity (like many connectionist models), but shall remain a mathematical /arithmetical fitting model as an alternative to connectionist simulation models. Based on these development goals, the NAMWR was developed by tackling the issues described above. In the following paragraphs I will explain how the refined model NAMWR deals with those issues.

Issue 1 (that the model could provide negative values) was solved by transforming the probability values into a normal distribution. If one takes a closer look at the data obtained by Nazir et al. (1992), one can see that a linear drop-off in recognition probability was obtained in a middle probability range while for very high and low probability values (towards zero and one) the curve seemed to flatten. While this flattening (especially towards zero) is not picked up by a linear decrease in recognition probability, it is picked up by the normal distribution function. The normal distribution function (like other distribution functions) is characterized by such a flattening towards high and low values while in the middle range it is fairly linear. Following the principle of nested modeling the linearity in the drop-off is kept, but it is transformed by the normal distribution function. The normal distribution function (rather than any other distribution function) was taken as a starting point because it is still the standard distribution function underlying most psychological and statistical models (e.g., the ANOVA or the MROM, Grainger & Jacobs,

1996, or the horse-race model of stop-signal inhibition, Logan & Cowan, 1984). Further evaluation of the model may well lead to the result that other distribution assumptions are more adequate (see Schmidt-Weigandt, 1999, for a review). However, the main point now is to start with a standard transformation that successfully prevents the model from yielding negative values.

$$p_{A,d}(i, r_f, y) = \begin{cases} \Phi(r_f - (i-y)d) & : \quad y \le i \\ \Phi(r_f - (y-i)Ad) & : \quad y > i \end{cases}$$

<u>Equation 9:</u> This is a normal transformation from Equation 4 with F being the standard normal distribution function of a random variable X ~ N(0,1). Rather than being a probability value, r_f has now been transformed into a z-value and recognition probability on fixation is computed by $F(r_f)$. As in Equation 4, A is the asymmetry ratio between drop-off rates, d the drop-off rate to the right, i the (ordinal) letter position, and y the spatial letter position. In this way, the new model can only produce positive values because $F(z) > 0$ for all z Î (¥, -¥).

The next step was to introduce a frequency parameter in such a way that frequency plays a greater role for poorer viewing positions and longer words:

$$p'_{A,d,l}(Freq, i, r_f, w, y) = p_{A,d}(i, r_f, y) + (1 - p_{A,d}(i, r_f, y)) \bullet \underbrace{l \bullet w \bullet (\Phi(\lg(Freq+1)) - 0.5)}_{=g_l(Freq,w)}$$

*<u>Equation 10:</u> This is the first real extension of the old model because a free lexical (or post-lexical) parameter is introduced. The recognition of each single letter is enhanced by (1- $p_{A,d}(i, r_f, y)$) * $g_l(Freq, w)$. l Î [0, ¥) is the lexical or post-lexical (cf. Paap & Johansen, 1994) parameter, Freq is the frequency per million, w the word length, lg is the logarithm to the base 10, and F is again the standard normal distribution function. The value of the resulting function $g_l(Freq, w)$ is thus a function the frequency Freq and word length w of a word that is modulated by the value of the free parameter l. The factor 1- $p_{A,d}(i, r_f, y)$ ensures that the influence of $g_l(Freq, w)$ gets greater for lower (prelexical) recognition probabilities.*

This simple extension catches all three frequency variations found in this thesis: The better performance for high-frequency words, the greater frequency effect for inferior viewing positions and the greater frequency effect for longer words in one or multiple ways. First, a lexical effect must be larger for higher frequencies. This is the case because the logarithm is a positively monotonous transformation of frequency. The higher the frequency of a given word is, the greater is the alteration by the frequency factor $g_l(Freq, w))$. Clearly, for nonwords, no frequency effect must be observed. This is the case as $g_l(Freq, w) = 0$ for $Freq = 0$. Thus, following the principle of nested modeling, in this respect the old model is contained in the new one. If no lexical influence (*Freq = 0)* is assumed, the model performs - except for the normal distribution transformation - like the old model without any lexical parameter *l*.

Second, the lexical effect should be larger for inferior viewing positions. This is the case as the influence of $g_l(Freq, w)$ is weighted by $(1 - p_i(y, A, d, r_f))$. Clearly, the smaller recognition probability $p_i(y, A, d, r_f)$ of the letter *i* is, the greater is the weight given to improvement of recognition $g_l(Freq, w))$.

Third, greater improvement for longer words is obtained in three ways. First, because the improvement recognition $g_l(Freq, w)$ is not computed depending on the word recognition probability, but depending on every single letter recognition probability, this improvement occurs more often in long words (because they contain more letters whose recognition probability can be improved). Second, as discussed and analyzed in Experiment 4, recognition for long words also deteriorates because letters far away from fixation are particularly badly recognized. Long words have more such letters. Recognition of those letters is particularly enhanced by $g_l(Freq, w)$ (see section above). Thus, in an indirect way, long words should also benefit more from this enhancement than short words. Third, and most obvious, lexical improvement is multiplied with word length. This may be related to the fact that even if one letter is badly recognized more information that may be lexically used is left in long words than in short words. This point is discussed in more detail below when an additional word length factor for the recognition on fixation is introduced.

Lexical improvement on a particular letter position *i* may be hindered when there are one or more neighbors on that letter position *i*. In Experiment 6, lexical information distribution as indexed by N-Grad altered the form of the viewing position curve and, consequently, the asymmetry parameter value in the fits with the old model (see problem 3). Performance generally improved when neighbors were on or near fixation compared to when they were off or far from fixation. In the revised model the original neighborhood distribution rather than N-Grad is introduced to account for this finding, thus allowing for an asymmetry parameter that is invariant to neighborhood distribution.

$$p''_{A,d,l}(Freq, i, posN, r_f, w, y) = p_{A,d}(i, r_f, y) + (1 - p_{A,d}(i, r_f, y)) \bullet \underbrace{\frac{g_l(Freq, w)}{0.5 + \frac{\Phi(posN_i)}{\overline{N}}}}_{h(i,posN)}$$

<u>Equation 11:</u> *In this formula, no new free parameter is introduced but rather the term h(i,posN). h is a function of the positional neighborhood distribution posN and the ordinal letter position i and corrects for the improvement by $g_l(Freq, w)$ of Equation 10. PosN is the vector of positional neighbors on each letter position and $posN_i$ is its ith element, i.e., the number of neighbors at the i'th letter position.* $\overline{N}$ *is the mean number of neighbors per position. The frequency effect is diminished when a high number of positional neighbors on letter position i are competing with the target for recognition.*

The model thus assumes that lexical processes do not help much to compensate for poor letter recognition when many neighbors are competing with the target word on that position. Thus, when $posN_i = 0$ (and $\overline{N}$ = 1) the improvement of letter recognition by $g_l(Freq, w)$ would be fully effective. However, earlier pilot fits showed that the inhibitory effect of *posN* is greater when only a few other neighbors compete than when many neighbors are competing on the other positions, too. This matches with assumptions in the MROM according to which one higher frequency neighbor tends to be particularly detrimental (Grainger & Jacobs, 1996). If more than one neighbor competes with the target word, they might also inhibit each other and hence reduce their inhibitory influence. The fewer neighbors a word has, the larger is the inhibitory influence of one particular neighbor in the model.

Note that in Experiment 6 positional neighbors had a particular inhibitory effect when they were far off fixation. In contrast, the absolute number of neighbors did not appear to have a general inhibitory effect, as the negative N-Grad conditions generally had more neighbors, but led to better performance. Usually, at least in the PIT neighbors tend to be inhibitory. Besides the hypothesis that they inhibit each other, there is another explanation for the results of Experiment 6. $\overline{N}$ might be a measure of a frequent sublexical phonological-orthographic structure which facilitates word recognition (Ziegler & Perry, 1998). Words with many neighbors tend to share more body neighbors. Moreover, because two of the three sublexical components, onset, nucleus and coda are usually shared by neighbors, the summed frequency of these sublexical components (SCF) tends to be higher for words with many neighbors. SCF has also been shown to facilitate visual word recognition (Nuerk, Rey, et al., 2000). Consequently, high $\overline{N}$ might decrease the inhibitory effects of $posN_i$. In the literature both inhibitory effects and facilitatory effects of neighborhood have been reported (see introduction; for a review, Andrews, 1997). The model accounts for the effects of Experiment 6 by assuming that any particular neighbor on a given position is inhibitory, while the whole (mean) number of neighbors per position might also represent a measure for a favorable sublexical orthographic-phonological structure. If future metrics for the facilitation of word recognition by favorable sublexical orthographic-phonological structure are developed and tested in future VPE experiments, $\overline{N}$ might be replaced by more refined and direct measures of sublexical structure.

Finally, shortcomings 2 and 5 must be tackled. The asymmetry parameter was originally thought to be a purely visual parameter, but varied as a function of word length. The perceptual frequency hypothesis assumes that this is not due to the visual asymmetry in recognizing letters (cf. Nazir et al., 1992), but due to the saccade landing position distribution in a word of a given length. This distribution is different in long

and short words (Nazir et al., 1998). Furthermore, the drop-off parameter varied between different word lengths. The revised model tries to have invariant visual asymmetry and drop-off parameters for different word lengths and tries to specify their alteration by word length explicitly.

$$\overline{p}_{A,d,r_f}(i,w,y) = \begin{cases} \Phi(r_f + 0.33(w-5) - (i-y)d) & : \quad y \le i \\ \underbrace{\Phi(r_f + 0.33(w-5)}_{\Phi(r_f,w)} - ((y-i)\underbrace{A(1+\frac{w-5}{10})}_{\overline{A(w)}}d) & : \quad y > i \end{cases}$$

*<u>Equation 12:</u> This formula corrects for word length influences on the visual parameters drop-off-rate d, and asymmetry ratio A. It does so by altering the asymmetry ratio and the recognition of fixation. First, as the asymmetry ratio is usually computed for this word length of five letters (see Experiments 1, 2, 4, 6, and, e.g., Nazir et al., 1998, for five- and nine-letter words)., the asymmetry ratio of five-letter words is taken as a standard. To this standard 0.1 is added for every additional letter and 0.1 subtracted for every letter less than five. In this way, the asymmetry of the viewing position curve may vary with word length, while the visual asymmetry parameter of single letter recognition may remain invariant. In addition, in Study 2 the drop-off rate had varied with word length: It was higher for shorter words. This was a compensation for the recognition at fixation being fixed to 1. Equation 12 deals with this problem in two ways. First, a word length correction of 0.33 * (w – 5) is introduced. Second, the recognition at fixation r_f itself is now a free parameter that can be fitted to the data. It may now vary separately from the drop-off rate. So, steep viewing position curves with a good recognition probability on fixation, but a high drop-off rate and rather flat curves with a low recognition probability on fixation, but a low drop-off rate become distinguishable.*

The drop-off parameter varied with word length and was much larger for short than for long words. This was due to a fixed letter recognition probability of 1 on fixation and only two letters (in three letter words) off fixation. These two letters were one or two letters away from the letter on fixation, so that a very high drop-off rate was needed to obtain a low word recognition probability for short words. The previous refinements had already favored long words in several ways: i) the incorporation of a lexical parameter was such that it helped more for long words, ii) recognition probability on fixation was allowed to vary so that the model could account for inferior performance in short words without assuming unreasonably high drop-off rates, and iii) a normal distribution transformation allowed greater drop-off rates for long words without an output of negative letter or word recognition probabilities. However, preliminary fits suggested that this was not enough to allow for an invariant drop-off rate. Recognition of short words was still overestimated and that of long words still underestimated. Therefore, I introduced the word length factor 0.33 * (w – 5) that enhanced the probability of recognizing a letter on fixation for

long words and decreased it for short words. As for the asymmetry ratio *A,* the visual parameter r_f remains invariant for different word lengths. However, the probabilities of recognizing the letter on and all the letters away from fixation correctly are increased by the word length factor. While the word length factor is necessary to fit the data with invariant parameters for different word lengths, it is not totally clear what it implies. However, a look at interactive activation models might help. In interactive activation models, all activated letters send activation to the lexical level, which in return sends activation back to the letters which are part of the activated lexical entries. The more letters there are in a word, the more activation accumulates. This poses the problem of simulating the word length effect successfully in some interactive activation models (Behrmann et al., 1998; Richter, personal communication), unless there is no word length corrections of some connection weight or other parameter (see Grainger & Jacobs, 1996, for such a correction which is quite similar to the one suggested here). In sum, any mathematical model borrowing the interactive activation idea has thus somehow to account for the fact that more activation from more letters is added up for long words than for short words. The increase in letter recognition probability in my model might reflect such an idea. If one or more letters in a long word are not recognized properly, more information is still available than in a short word. However, other accounts, such as different sublexical structures for different word lengths, are also possible. My model, like any mathematical model, is a combination of quantitatively testable hypotheses of mathematical structure. The incorporation of the new free parameters and the fixed parameter values are based on certain conceptions about word recognition. The specific fixed parameter values of the recognition of fixation being 0.33 * (w – 5) and of the asymmetry variation being 0.1 were chosen to be capable of successfully fitting the data. Clearly, these alternative ideas and their parametrical incorporation should be systematically tested in future experiments.

So putting together the above equations, we get the final equation for the probability to fixate a letter at the ordinal position *i.*

$$q_{A,\,d,\,l,\,r_f}(Freq, i, posN, w, y) = \overline{p}_{A,\,d,\,r_f}(i, w, y) + (1 - \overline{p}_{A,\,d,\,r_f}(i, w, y)) \bullet \frac{g_l(Freq, w)}{h(i, posN)}$$

<u>Equation 13:</u> Equation 13 is essentially the same as Equation 11. However, the (lexically or post-lexically and orthographically) uncorrected letter recognition probability $p_{A,d}(i, r_f, y)$ has changed from Equation 11. $\overline{p}_{A,\,d,\,r_f}(i, w, y)$ *is the letter recognition probability which is corrected for parametrical invariance deviations caused by word length (see Equation 12).*

$$QW_{A,d,l,r_f}(Freq, posN, w, y) = \prod_{i=1}^{w} q_{A,d,l,r_f}(Freq, i, posN, w, y)$$

Equation 14: Following the principle of nested modeling, the independence assumption that the probability of recognizing a word is the product probability of recognizing its constituent letters is still assumed to be valid. Thus, the probability of recognizing the whole word is analogous to Equation 5: It is the product of the probabilities recognizing the single letters i = 1, ..., w. The function arguments in parentheses Freq, posN, w, y are then the arguments that are given by stimulus material and presentation conditions, while the free parameters A, d, l, r_f define the parametric family of viewing position curves. The parameter combination with the most adequate fit is selected from this parametric family. In the same way as in Equations 6 and 7, y could be substituted as given in Equation 2.[17]

To summarize, the following steps have been taken to provide a possible solution to the problems named in the introduction of this paragraph:

1. The model could produce negative word recognition values.

A normal distribution transformation of the linear drop-off was introduced to prevent this.

2. The asymmetry parameter (originally interpreted as a purely visual parameter) changed as a function of word length.

A word length influence on the asymmetry of the viewing position curve was explicitly specified and no longer included in the visual asymmetry parameter.

3. The asymmetry parameter (originally interpreted as a purely visual parameter) changed as a function of neighborhood distribution.

The influence of neighborhood distribution on the asymmetry of the viewing position curve was explicitly specified and no longer included in the visual asymmetry parameter.

4. The visual drop-off parameter changed as a function of frequency.

A lexical parameter was introduced that improves letter recognition of each single letter dependent on frequency of the target word. Thus, frequency differences can be fitted with an invariant drop-off parameter.

5. The visual drop-off parameter changed as a function of word length.

Five things directly and indirectly improved fits for long words to keep the drop-off parameter invariant for future fits. First, in multiple ways, frequency helped more for longer words. Second, it helped more for

[17] The program implementations and an overview over all model parameters can be found in Appendices J,K, and L.

inferior letter recognition probabilities more often occurring in long words. Third, normal distribution transformation allowed steeper drop-off rates for long words without resulting in negative word recognition values. Fourth, the probability of recognizing letters on fixation was allowed to vary to make smaller drop-off rates for short words possible. Fifth, a word length influence on this probability was incorporated.

6. For long words, the fitted curves were too flat.

The incorporation of a lexical parameter in addition to the drop-off rate, the variation of recognition probability on fixation, and the influence of word length on recognition probability on fixation allowed for steeper word recognition curves.

7. The ratio of fitting five data points with two free parameters is already not a good ratio – adding further parameters only makes sense when the number of data points fitted is increased.

All improvements mentioned above allowed the fitting of curves for different word lengths, frequencies and neighborhood distributions in one experiment with the same invariant parameter set.

8. The probability of letter recognition on fixation is fixed to 1.

By allowing the model to fit 30 data points simultaneously, another free parameter is possible while still keeping a better ratio of free parameters to data points than before.

While the model for the first time now allows different curves from one experiment to be fitted with the same invariant parameter set, it does so by using less free parameters. While in Experiments 4 and 6, 12 free parameters were used to fit 30 data points (2 free parameters for five data points of each curve), now this number is reduced to four free parameters. Will the NAMWR with this reduction still be able to fit the data as successfully or even more successfully than the old model?

5.4. Fitting the Experimental Results with the NAMWR

The fits of the NAMWR were, like those of the old model, performed with the Matlab Simplex algorithm. When I used other fitting algorithms (e.g., Levenberg-Marquardt), I obtained almost the same results (where “almost the same” means differences of approximately 1/1000 to 1/1000000 for any parameter, but no more). Thus, the optimal solutions seem to be clear in the current experiments. Figures 43-45 and Table 14 show that fitting 30 data points with four free parameters provided satisfactory results (all RMSDs < 0.05 for p-values $\in [0, 1]$). Let me first consider the fittings of the results of Experiment 4 in greater detail (see Figure 43 and Table 14).

Fitting Parameters and Fitted Values for All Experiments and All Viewing Positions											
#Letters	**Freq**	RMSD	d	A	l	$\Phi(r_f, w)$	VP1	VP2	VP3	VP4	VP5
3	**low F**	0.046	0.500	1.26	0.35	0.93	0.49	0.60	0.65	0.60	0.49
3	**high F**	0.046	0.500	1.26	0.35	0.93	0.65	0.73	0.76	0.72	0.64
5	**low F**	0.046	0.500	1.26	0.35	0.98	0.36	0.58	0.63	0.44	0.18
5	**high F**	0.046	0.500	1.26	0.35	0.98	0.60	0.75	0.78	0.66	0.45
7	**low F**	0.046	0.500	1.26	0.35	1.00	0.24	0.56	0.55	0.16	0.01
7	**high F**	0.046	0.500	1.26	0.35	1.00	0.58	0.78	0.75	0.48	0.23
N-Grad	Freq	RMSD	d	A	l	$\Phi(r_f, w)$	VP1	VP2	VP3	VP4	VP5
N-Grad < 0	**low F**	0.049	0.408	1.23	0.29	0.97	0.48	0.64	0.67	0.54	0.33
N-Grad < 0	**middle F**	0.049	0.408	1.23	0.29	0.97	0.68	0.78	0.78	0.68	0.51
N-Grad < 0	**high F**	0.049	0.408	1.23	0.29	0.97	0.81	0.87	0.86	0.78	0.64
N-Grad > 0	**low F**	0.049	0.408	1.23	0.29	0.97	0.41	0.58	0.63	0.51	0.29
N-Grad > 0	**middle F**	0.049	0.408	1.23	0.29	0.97	0.53	0.68	0.73	0.64	0.46
N-Grad > 0	**high F**	0.049	0.408	1.23	0.29	0.97	0.53	0.68	0.73	0.66	0.48
Synt. Class	Form	RMSD	d	A	l	$\Phi(r_f, w)$	VP1	VP2	VP3	VP4	VP5
Nouns	**ALC**	0.020	0.415	1.08	0.31	0.96	0.50	0.63	0.67	0.58	0.42
Nouns	**IUC**	0.021	0.239	1.40	0.31	0.96	0.70	0.76	0.76	0.69	0.56
Nouns	**AUC**	0.028	0.361	1.14	0.31	0.96	0.56	0.68	0.70	0.62	0.46
Non-Nouns	**ALC**	0.025	0.282	1.45	0.28	0.98	0.73	0.80	0.80	0.72	0.55
Non-Nouns	**IUC**	0.028	0.263	1.55	0.28	0.98	0.75	0.81	0.81	0.72	0.56
Non-Nouns	**AUC**	0.017	0.408	1.08	0.28	0.98	0.56	0.71	0.76	0.68	0.51

Table 14: Fitting Parameters, deviations and fitted accuracy for Experiment 4 (top), Experiment 6 (middle) and Experiments 1 and 2 (bottom). RMSD: Root mean square deviation, d: Drop-off parameter to the right, A: Asymmetry ratio, l: Lexical parameter; $F(r_f, w)$: Recognition probability at fixation, VP: Viewing positions within a word, N-Grad: Neighborhood gradient, Freq: Frequency; Synt. Class: Syntactic Class; Form: Word Form. Note that in Studies 2 and 3 all different curves are fitted with an identical set of four free fitting parameters. Thus, the differences are only due to stimulus properties.

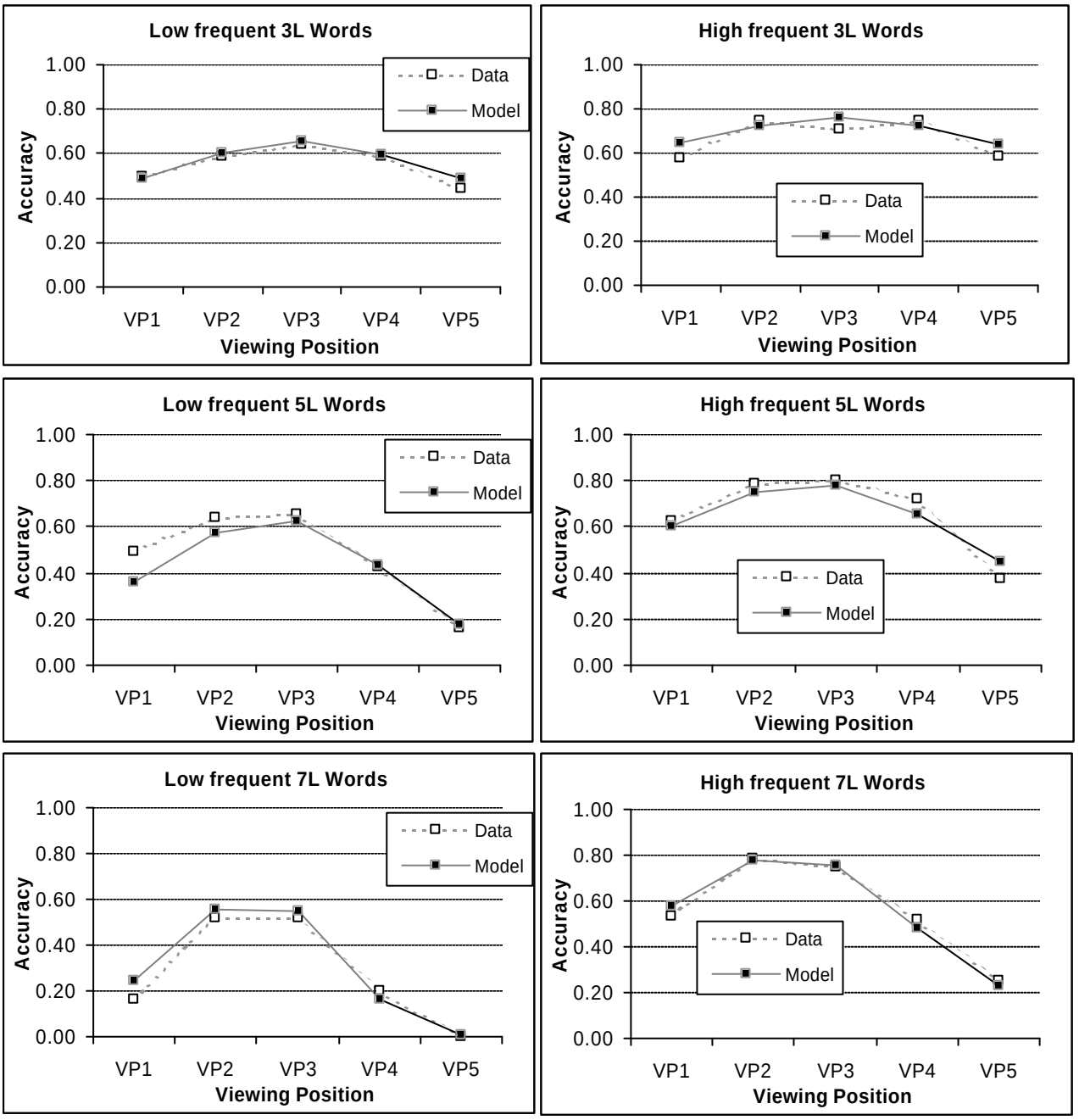


Figure 43: *Empirical and fitted curves for different word lengths and frequencies. All six curves are fitted with an identical parameter set. Note that, in particular for long and short words, an identical drop-off rate is now used. The fitted curves for long words were no longer too flat as they had been for the old model. However, the NAMWR did not capture the M-shaped curve for high-frequent three-letter words. If such curves were reliably produced, the model would be falsified.*

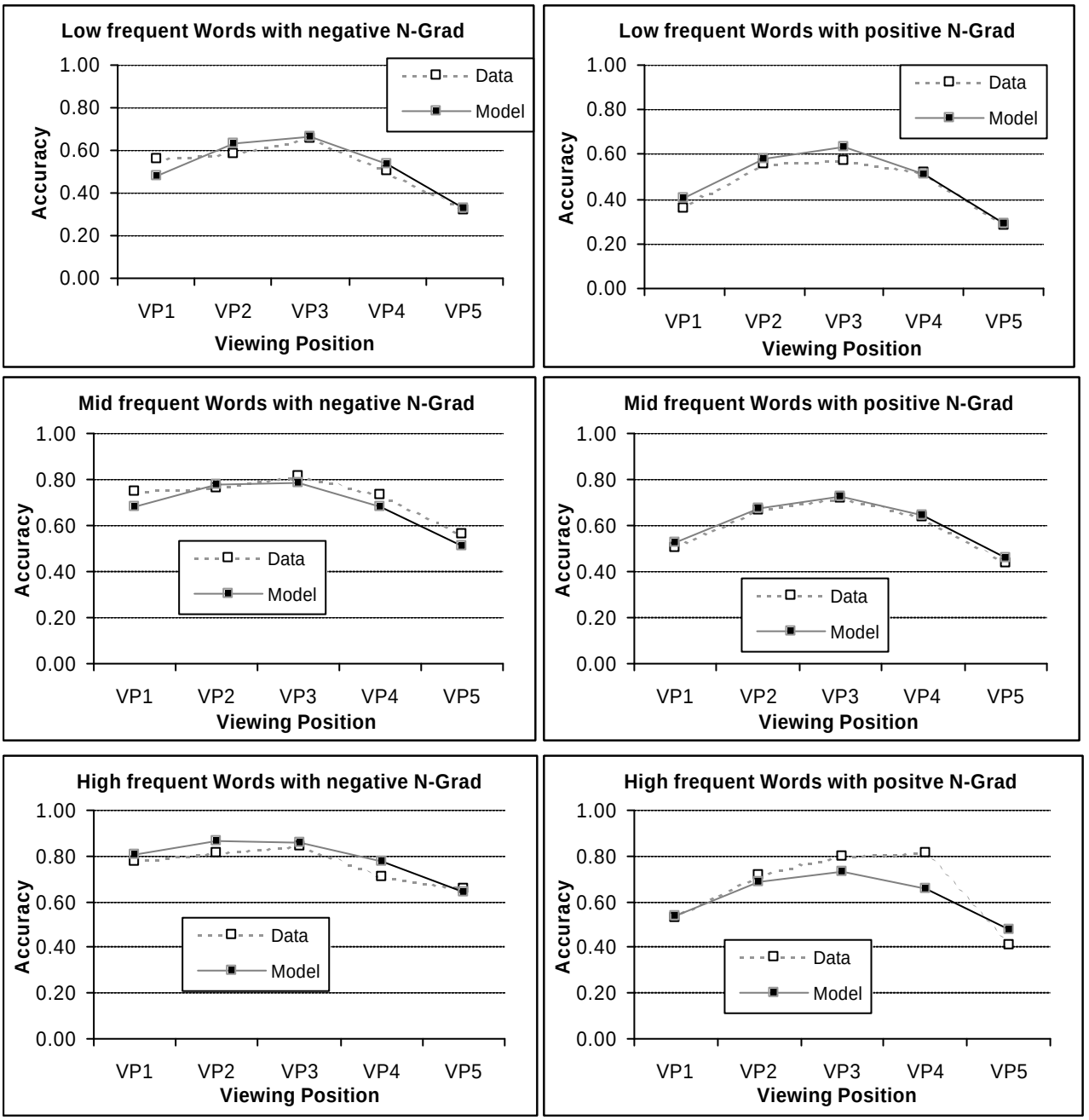


Figure 44: Empirical and fitted curves for different N-Grads and frequencies. All six curves were fitted with an identical parameter set. Note that the differences in asymmetry were successfully captured by the model with an identical asymmetry parameter, simply taking into account neighborhood distribution (consider the first and final viewing position for symmetry evaluation). For high-frequent words with negative N-Grad, the model did not well capture the unexpected high drop-off from viewing position 4 to viewing position 5.

The second aspect is particularly important because this is a falsificator of the model. If in the absence of any extreme variation of a particular linguistic, non-visual variable (such as positional neighborhood) a decrease in performance from an outer curve point to an inner curve point with subsequent increase is obtained (a chainsaw M-shape), then the model is architecturally wrong. Such a decrease may be obtained, for example, for special neighborhood distributions of words that have all their neighbors at the

beginning. Then performance on the first position could be optimal and decrease towards the end of the word. However, it must not increase again on the final viewing positions. The only circumstances in which this is possible are very extreme variations of neighborhood distributions. If all neighbors are either on the first or on the last letter of a long word, the model may produce relatively better performance on the respective fixation positions even though they are on the edge of the word. In the absence of any such extreme manipulation, an M-shape must not be observed. If such results are reliably obtained, one ore more basic assumptions of the NAMWR would be wrong. Hence, I hypothesize that the M-shaped curve is due to random statistical variance in the VPE curve for high-frequent three-letter words, i.e., that the "true curve" might show a higher performance on the middle position. I do not know of any study that has reliably obtained an M-shaped viewing position curve. With this exception, the fits of the results of Experiment 4 were successful. I will now discuss the fits of the results of Experiment 6, in which frequency, neighborhood distribution, viewing position, and their interaction with one another were investigated (see Figure 44 and Table 14).

The model captured the variation of the asymmetry of the viewing position curve with N-Grad well (RMSD = 0.049 < 0.05), although the visual asymmetry parameter was invariant for all curves. While all of the fitted curves for positive N-Grad were fairly symmetrical, the fitted curves for negative N-Grad were not (compare performance on positions 1 and 5 for each curve.) The model again fitted the frequency effect and its interaction with the VPEs well: Poorer viewing positions showed a greater benefit from high-frequent stimulus material. Note that in this experiment there were three monotonous frequency groups rather than two as in Experiment 4. The model did not only generally capture the frequency effect for high- vs. low-frequent words, but also the quantity of the continuous increase in performance from low- over middle- to high-frequency items. However, two effects were not fitted well: First, the curve for high-frequent five letter words with positive N-Grad with an unusually high drop-off from position 4 to position 5 could not be picked up by the model. However, this pattern of results already concerned us in the discussion of Experiment 6. Descriptively, the viewing position curve for the 5th position was lower for high-frequency words than for middle-frequency words, while for the 4th position it was, as expected, vice versa. Because similar neighborhood distributions were selected in these two stimulus groups, the model cannot predict such a crossover of similar curves in both groups. It could only predict such a crossover between stimulus groups if neighborhood distributions differed. Empirical replications are needed to find out whether or not this crossover is reliable and systematic or a noise artifact. Finally, the model predicted a somewhat greater decrease in performance for the low- and middle-frequency group with negative N-

Grad than actually found in the data. This is in contrast to the four other groups. The finding may point to a possible over-simplification in the assumption that all neighbors are equally inhibitory (when viewing position distance of fixation and neighborhood position is kept constant). However, the inhibitory effects of different neighbors may differ. Recall the discussion of the results of Experiment 6 in Study 3. Neighbors at the beginning may be Body Neighbors (cf. Ziegler & Perry, 1998). Recognizing them well (i.e., when fixating on a word's beginning) may improve performance particularly in low- and middle-frequent words because frequent sublexical orthographic-phonological structures are known to be especially effective in low-frequency words (Nuerk, Rey, et al., 2000). If neighbors on the end do not activate such sublexical orthographic-phonological structures, performance may be less improved. In sum, if not all positional neighbors are equally influential in orthographic-phonological-lexical space, this pattern of results could be potentially expected. It could not yet be fitted by the NAMWR because phonology has not been incorporated yet. Therefore, not capturing those effects does not falsify the model because orthographic-phonological effects are beyond its current scope.

Finally, a comparison of the parameter values of Experiments 4 and 6 is very instructive for examining the quality of the model. It has long been known that there are individual differences even between normal participants in visual word recognition performance (e.g., Butler & Hains, 1979). Thus, the parameters may vary somewhat between the two experiments because different participants were tested. However, it is astonishing that the parameters remained quite invariant over the two experiments. Recognition on fixation and visual asymmetry ratio were almost identical in the two experiments and drop-off rate[18] (0.41 vs. 0.50 standard deviations) and lexical parameters (0.29 vs. 0.35) were still in a similar range. However, in Experiment 4, long words were part of the stimulus material. In those words strong lexical top-down processing would be of greater use because of their lower (mean) number of neighbors. For shorter words with a higher (mean) number of neighbors, lexical top-down processing could easily lead to a selection of a wrong candidate from the lexicon. Does the variation in the lexical parameter indicate that participants differed somewhat in their strategy in the two experiments or is it merely an artifact of the model? Here again, the model leads to an interesting question for future studies. Can lexical processing be strategically manipulated in a PIT? Grainger and Jacobs (1996), for example, assumed a strategic process for the LDT, but not for the PIT. If a strategic variation is assumed in a particular experiment, then the lexical parameter in the NAMWR should be consistently higher under conditions which favor the heavy use of lexical top-down processes (e.g., in words with few or no neighbors as

[18] Recall that the drop-off rate is a z-value in the NAMWR rather than a p-value as in the old model.

compared to words with many neighbors, or in pure word lists compared to mixed word and nonword lists.) Using the NAMWR, allows one to test this hypothesis by fitting the lexical parameter in different experiments and examining its values. If Grainger and Jacobs (1996) are right, this lexical parameter should not systematically vary in a PIT under those different experimental conditions.

Fitting the results of Experiments 1 and 2 with the current model is not possible using invariant parameters. The conditions in each experiment do not differ in any fixed parameter incorporated in the model because the words presented in IUC, ALC and AUC form were identical in length, frequency and neighborhood distribution. A perceptual frequency hypothesis of visual word form has - as introduced before - not been investigated in the current form. In particular, no one has ever developed and systematically tested a continuous metric of perceptual frequency, as was done for lexical frequency. For lexical frequency, it was found in different tasks that the logarithm of frequency was a better index of lexical processing than raw frequency (e.g., Schmidt-Weigand, 1999; Ziegler et al., 1998). No such evaluation was ever performed for perceptual frequency. We do not yet know anything about the relationship of frequency-dependent case-specific and abstract letter processing in any word recognition task. Thus, because no metric and no task-overlapping systematic evaluation of the learning curve of perceptual frequency of item-specific visual word form exist, no model has yet incorporated any parameter for perceptual frequency of visual word form. In my opinion, there are simply not enough constraints yet to implement anything meaningful. Therefore, it is difficult to compare new fits with old fits. However, what can be done, is the following: The mean performance in Experiments 1 and 2 can be fitted. Based on these results the two new free parameters of the model (lexical parameter and recognition probability on fixation) can be fixed. Then, asymmetry ratio and drop-off rate can be fitted as was previously done in the old model for each single curve of each experiment. The results can be seen in Figure 45 and Table 14: As regards nouns, the fits were good and particularly for ALC and AUC form somewhat better than before. It apparently helped in these case forms that lexical attributes of the stimuli, the normal transformation of the drop-off rate, and a variable recognition probability on fixation were considered in the NAMWR. For non-nouns, I obtained a similar picture. As concerns AUC form, markedly better fits were observed than before, while for the better IUC and ALC conditions the fits did not differ much. In sum, the new NAMWR fits were better than or at least equal to those of the old model.

Altogether, the fits of the NAMWR are quite successful and interesting because further research hypotheses can be specified and quantitatively tested using the NAMWRs constraints and predictions. Thus, on the basis of these data, the NAMWR seems a promising step towards a general mathematical

model of visual word recognition. In the next paragraph, I will discuss how the model refinement fares with respect to modeling principles and criteria.

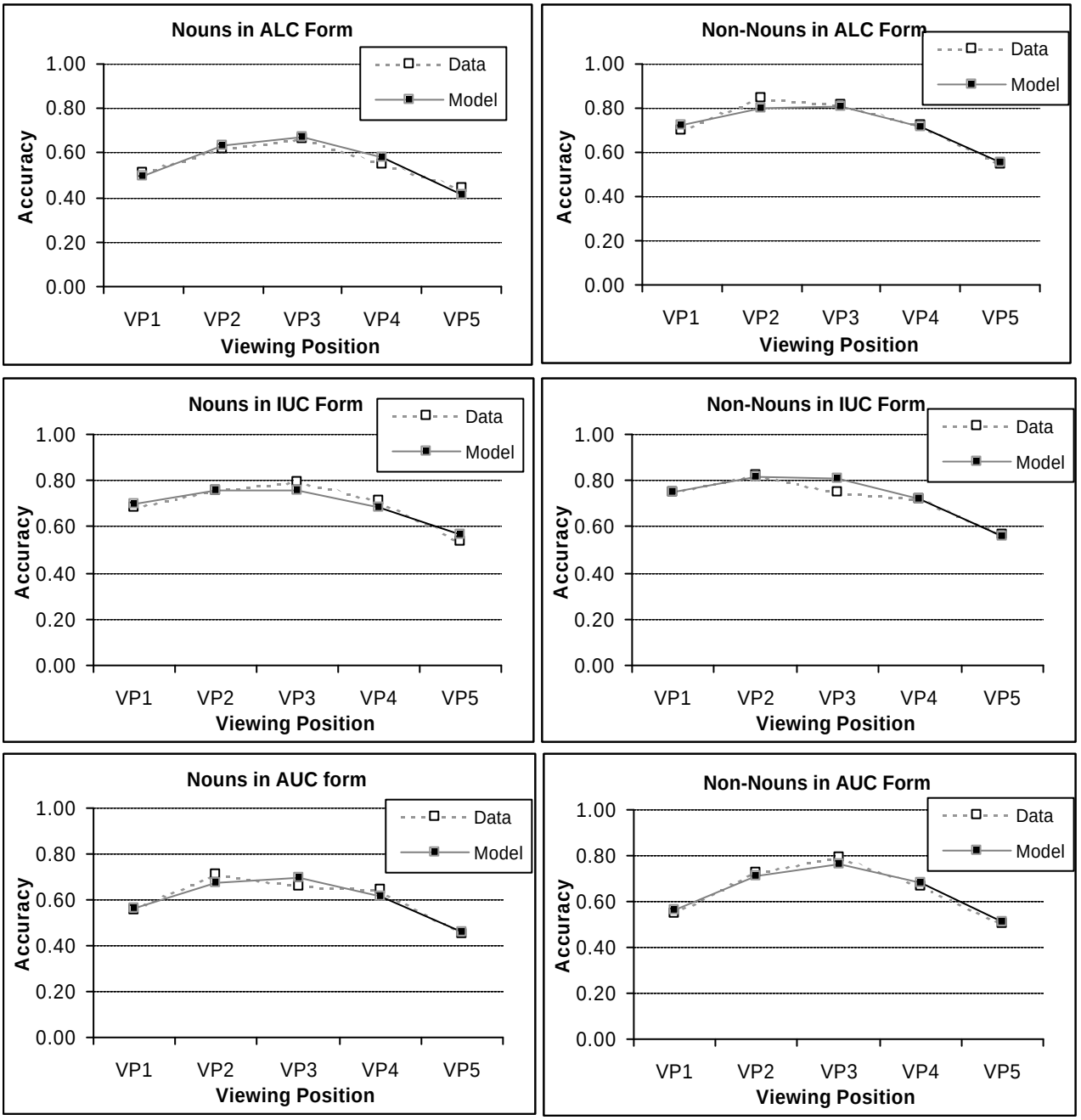


Figure 45: Empirical and fitted curves with the NAMWR for different syntactic classes and word forms. Because no learning curve yet exists for the perceptual frequency of visual word form, it was not possible to introduce such a parameter. Therefore, the lexical parameter and the recognition at fixation were fixed across the stimuli for each experiment. The single curves were then fitted by varying drop-off rate and asymmetry ratio as in the old model. The fits with the functional form of the NAMWR were better for non-nouns (RMSD = 0.024 for the whole experiment for the NAMWR vs. 0.035 before) and about equal for nouns (RMSD = 0.024 for the NAMWR and RMSD = 0.022 before) compared to the old model.

5.5. The NAMWR in the Light of Current Modeling Criteria and Principles[19]

5.5.1. Introduction

"A measure of advancement in psychology [...] is the discovery of general laws and principles that govern the phenomenon under investigation. [...] How should on decide among a set of competing models for the same phenomenon? This model selection problem is the heart of scientific enterprise."

(Myung, Forster, & Browne, 2000, p. 1)

Successfully modeling the data does not imply that a model is appropriate. In modeling research multiple selection criteria, design and system principles have been discussed to decide which model is (most) appropriate to fit the data. Extensive reviews and discussions of these principles and criteria can be found in Forster (2000), Jacobs and Grainger (1994), Jacobs et al. (1998), Myung (2000), Myung and Pitt (1997), Ratcliff (1988), Van Orden and Goldinger (1994), Van Zandt and Ratcliff, (1995). I will discuss the NAMWR with respect to the selection criteria suggested by Jacobs and Grainger (1994), which are most often considered as relevant in the literature (e.g., Myung & Pitt, 1997): Descriptive adequacy, explanatory adequacy, simplicity/ falsifiability (complexity), and generality. I will discuss the simplicity and falsifiability of the model separately because although simplicity can be linked to falsifiability (Jacobs & Grainger, 1994; Popper, 1935), those two criteria are not necessarily identical in all aspects (Jacobs & Grainger, 1994; Myung & Pitt, 1997). Finally, I will discuss how the NAMWR fares with respect to the design principle of nested modeling. Naturally, these criteria and/or principles are just a selection of the multiple criteria or principles proposed by different authors. An undisputed canonical set of criteria / principles has not yet emerged and the discussion which criteria / principles are necessary or most relevant can be found in the above literature, but is beyond the scope of this thesis. However, those six criteria / principles create a starting point for model evaluation (Jacobs & Grainger, 1994; Myung & Pitt, 1997).

5.5.2. Descriptive Adequacy

The selection criterion of descriptive adequacy is considered the "foremost criterion of model selection" (Myung & Pitt, 1997, p. 80). Descriptive adequacy implies that a model is adequate only if it

[19] Some modeling evaluation aspects are called criterion in the literature (e.g., descriptive adequacy criterion) while others are called principle (e.g., nested modeling principle). It may well be questioned whether this distinction makes much sense. However, the discussion which model criteria /principles are the most important for model evaluation and how they should be named is beyond the scope of this thesis. Therefore, the evaluation aspects are called principles and / or criteria as in the literature.

provides a good fit to the data. The fit is estimated by measures such as i) percent of variance accounted for by the model in a regression analysis, or error measures of deviation such as ii) the root mean square deviation (RMSD), or iii) the mean square error (MSE) between or within groups used in the ANOVA for the inference statistics with the F-distribution. The criterion of descriptive adequacy is a *necessary* criterion because a model that does not fit the data pattern sufficiently accurately cannot be considered to be a close approximation to the truth. Note that it is not a sufficient criterion because an infinite number of models can be found which fit a limited number n of data points perfectly (such as polynoms of dimensions n+1 and greater).

While in inference-statistical models the scientific community has developed some standards (alpha = .0.95), in non-inferential statistics no such standard has yet emerged (cf. Collyer, 1985; Myung & Pitt, 1997). For example, Myung and Pitt (1997) criticize Massaro and Cohen (1993) for setting an arbitrary limit on the interpretation of goodness of fit measures in defending their FLMP model against criticism from Cutting, Bruno, Brady and Moore (1992). Still, there is a need to quantify the meaning of 'the model fits the data well'. In this thesis and the following discussion I follow the conventions for inference statistics in my interpretation of probability data. An RMSD < .05 is considered a good fit and an RMSD > .10 a non- adequate fit to the data. The NAMWR fits 30 data points with a RMSDs < .05 using four free parameters and, hence, is considered descriptively adequate.

However, as noted above these good fits are necessary for the NAMWR to be considered relevant, but they are by no means sufficient. The best fit does not naturally imply that the respective model is the best or the true one (Myung & Pitt, 1997; Schmidt-Weigand, 1999). The true model may not be included in the tested set. However, even if it is included, it may not provide the best fit because it may provide the best fit for data minus noise, but not for data plus noise. Moreover, for a limited data set, the underlying model or distribution may not be the most appropriate (e.g., take the t-distribution for a limited set of standard normal distributed random variables which only converges to the normal distribution as the set size goes towards infinity). Finally, and most importantly for any given finite set of n data points, a model that perfectly fits the data can always be found: It is the polynom of degree n-1. However, this model rarely adheres to the following two principles: Explanatory adequacy and complexity / simplicity.

5.5.3. Explanatory Adequacy

Because one can always fit any data set perfectly, any fitting model is by itself nothing more than statistical mimicking (Ratcliff, 1988; Van Zandt & Ratcliff, 1995). Put more bluntly, "no functional relation

that is found to describe a pattern of data is in itself (psychologically) meaningful." (Schmidt-Weigand, 1999, p. 37). Explanatory adequacy, however, has not yet been quantified as a criterion, nor is it clearly defined. Grainger and Jacobs (1994) propose that a model should have a minimum of ad-hoc assumptions. Myung and Pitt (1997) do not even discuss this criterion. For this thesis, I will focus on the meaningfulness of the parameters in my model - that their variation is not ad hoc, but clearly interpretable based on psychological empirical findings and theory.

The NAMWR in its current form was designed to comply to the principle of explanatory adequacy in that the psychological meanings of the parameters in the model are clearly defined and separated from each other. This psychological adequacy was tested and can further be tested in future experiments. The drop-off parameter is the drop-off of letter legibility with distance from fixation. It has been shown that letter legibility systematically drops off with distance from fixation (Nazir et al., 1992) and this drop-off is asymmetrical (Nazir, 2000; Nazir et al., 1992). Under invariant visual conditions, thus, the drop-off rate and asymmetry ratio should be invariant. In particular, different viewing position curves for different stimuli should be successfully fitted with identical parameters. This is the case for the six different conditions of Experiments 4 and 6, but was not the case before the model refinement. In this respect the NAMWR is a real step forward. In a literal way "the limitations [of the previous model provided] the impetus for the next generation of research" (Seidenberg & Plaut, 1998, p. 236).

However, with respect to explanatory adequacy the limitations of the current model can also clearly be seen. There exists no parametrical account for perceptual frequency effects of word form. If the model fitted different curves to the empirical data for different word forms, they would have to be fitted with varying drop-off and asymmetry parameters that were not originally thought to code visual word form. However, no current model can adequately account for perceptual word form effects because there is neither a case-specific letter representation nor a perceptual frequency metric. Models would need some parametric perceptual frequency parameter for perceptual word form (as, for example, case-specific letter nodes with particular connection strengths to different perceptual lexicons in an interactive activation framework). No model has such nodes or connections or can even distinguish between the manipulations in Experiments 1 - 3. All current models of visual word recognition including the NAMWR are therefore not adequately explanatory in this respect. The impetus for the next generation of research is already at the horizon here. Overcoming these limitations may lead to future research that not only establishes the perceptual word form effect as such, but also its learning curves for different letter and word forms.

5.5.4. Simplicity (Model Complexity)

"Models, as people have known since the days of Copernicus, have to be as simple as possible" (Jacobs & Grainger, 1994, p. 1317). The problem, however, is which of two different models is the simpler one? How are simplicity and complexity measured and defined? Myung and Pitt (1997) extended earlier considerations in pointing out that the increase in complexity may not be examined along a single dimension. Instead, they suggested to distinguish at least three dimensions of a model's complexity: Number of parameters, functional form and extension of parameter space. I will discuss the complexity of the NAMWR with respect to these three dimensions.

Number of Parameters

I have already pointed out that a polynom with degree n-1 can always perfectly fit a curve with n data points. Usually, the more free parameters a model has, the better its fit to a fixed number of data points is. Vice versa, the more data points have to be fitted, the worse performance with a fixed number of parameters is. In the NAMWR, the number of free parameters was increased from two to four compared to the model of Nazir et al. (1998). However, the NAMWR was capable of fitting 30 data points in Experiments 4 and 6 with four free parameters, while the old model needed 6 * 2 = 12 free parameters to fit the six curves with five data points each. While a reduction from 12 free parameters to four parameters is generally advantageous, one needs to relate the goodness of fit of the old model to that of the new model. As a measure of fit, I had computed the root mean square deviation which is the root of the mean of summed squared errors (SSE). For this case, Myung and Pitt (1997) suggested the approximation formula RMSEA[20] (Steiger, 1990) which is defined as:

$$RMSEA = \sqrt{\frac{F_i}{N - n_i}}$$

Equation 15: F(i) is the measure of the lack of fit of the model i (in this case the SSE), N the Number of data points, and n_i the number of free parameters. The formula penalizes complex models by subtracting the number of free parameters from the data points.

If RMSEA is computed for Experiments 4 and 6, the old model fits the 30 data points worse than the new model ($RMSEA_{old}$ = 0.93 > 0.050 = $RMSEA_{NAMWR}$ in Experiment 4 and $RMSEA_{old}$ = 0.063 > 0.053 =

[20] Note that the RMSEA is not an absolute measure of deviation between model and data, but a measure that enables a comparison of different models with different parameters. The RMSEA is not very sensitive to the number parameters if the number of data points is much greater than the number of parameters because the latter is subtracted from the first. Thus, the formula favors the introduction of an additional parameter for a slightly better fit. If a division rather than a subtraction were used, the approximation would favor models with a small number of free parameters more than the RMSEA does now.

$RMSEA_{NAMWR}$ in Experiment 6). Thus, the NAMWR needs far less parameters for fitting 30 data points than the old model. Given this reduction in the number of parameters, it provides a better fit than the old model in both experiments. This is remarkable because the RMSEA is not very sensitive to the reduction of free parameters when the number of data points is large. The number of parameters used to fit the six curves was reduced by factor 3 (from 12 to four). If such a relative reduction were considered, rather than the difference from a much higher number of data points, the improvement of the model would even be stronger than it is now. However, for Experiments 1 and 2 the NAMWR fares about as wells as the old model because each curve still had to be fitted by adjusting drop-off rate and asymmetry ratio (no parameter of perceptual frequency is yet introduced). For Experiment 1 the RMSEAs are quite similar, while for Experiment 2 the new model fared somewhat better ($RMSEA_{old} = 0.029 < 0.030 = RMSEA_{NAMWR}$ in Experiment 1 and $RMSEA_{old} = 0.046 > 0.030 = RMSEA_{NAMWR}$ in Experiment 2). Thus, with respect to the number of free parameters, the NAMWR is simpler than the old one because the NAMWR uses fewer parameters while maintaining a reasonable goodness of fit.

Functional Form

While the number of parameters is a well known and fairly accepted measure of simplicity, functional form is *"an often unrecognized dimension of model complexity that significantly affects model fits by simply capturing irrelevant patterns of the data"* (Myung & Pitt, 1997, p. 81). Myung and Pitt (1997, p. 81) define functional form *"as the way in which parameters are combined in the model equation".* They illustrate the importance of functional form by discussing the differences between the Fuzzy Logic Model of Perception (FLMP) of Massaro and colleagues (Massaro & Cohen, 1993; 1994; Oden & Massaro, 1978) and the Linear Integration Model (LIM) of Anderson (1981). With respect to simplicity as measured by the number of parameters the FLMP and LIM are equally complex. However, with respect to functional form the models seem to differ greatly. Myung and Pitt (1997) claim that the functional form of the FLMP is much more complex because it fits any empirical or simulated data pattern better than the LIM – the FLMP even fitted data patterns produced by the LIM better than the LIM itself (see also Cutting et al., 1992; Massaro & Cohen, 1993). So, not only the number of parameters determines how well a model can fit the data, but also the way in which these parameters are combined. A flexible combination of parameters may enable a model to fit a larger number of data patterns with a smaller number of parameters. A less flexible functional form allows only a limited set of curve shapes and more specific predictions for future data shapes. Note that functional form is often linked with the idea of falsificationism (cf. Grainger & Jacobs, 1996; Jacobs & Grainger, 1994). Strong falsification mostly implies producing an

effect (i.e., a pattern of data) that the model is unable to handle with its current functional form (i.e., the patterns of data it is able to produce). This is also the case for the falsificator of the NAMWR.

Like the old model of Nazir et al. (1998), the NAMWR is very restricted in its functional form. Except for possible extreme neighborhood distributions, the model will always produce concave curves with a drop-off to both sides from optimal fixation, and convex shapes only when the curve approaches zero. Other forms like strong chain-saw curves cannot be produced by the model. Moreover, the new model is more restricted in its functional form than the old model in that it makes specific predictions as to how the asymmetry of the curve changes as a function of neighborhood distribution. Performance will always be relatively better when fixation is on or near letter positions that are particularly critical for neighborhood letter exchanges. Finally, the new model also makes specific predictions about the quantitative pattern of frequency effects and about how they interact with visual variables and word length. The functional form of the model is such that under inferior visual conditions the frequency effect should be stronger than under near-optimal viewing conditions. In this respect, the model is very simple because the variability of functional form is very restricted and provides specific falsifiable hypothesis for future experiments.

Extension in Parameter Space

The third complexity criterion suggested by Myung and Pitt (1997) is the extension in parameter space. The parameter space refers to the range in which parameters may vary. The idea is that a smaller range of parameters implies that a model is simpler. Myung and Pitt (1997) give the example that a model that allows a parameter to vary in [- R, R] is more complex than a model that allows a parameter to vary in [0, R]. In the NAMWR the parameters usually extend only over a small range. The drop-off rate, the asymmetry ratio, and also the lexical parameter must be positive. In the program this is implemented by calculating the absolute values of these parameters in the fitting procedure. Only the free parameter that codes recognition likelihood on fixation can, in principle, vary from positive to negative infinity, but it is transformed via the normal distribution so that its respective recognition values are between 0 and 1.

The latter point illustrates a general difficulty of the third criterion of Myung and Pitt (1997). Transformations of parameters can be used to restrict the range in which these parameters vary and create seemingly simple models. Consider the following example: I may restrict a parameter value in (0, 1). By taking this value and computing its inverse normal distribution z-value this value ranges from ($-\infty$, ∞). For restricted ranges, the problem is even more straightforward. Any restricted range of a parameter is by a simple linear transformation $f(x) = ax+b$ equivalent to any other. In particular, a parameter φ

ranging from [0, R] would be absolutely equivalent to a parameter $\lambda \in$ [-R, R], if the model computed the linear transformation. $\varphi = f(\lambda) = \lambda/2 + R/2$. With this linear transformation the resulting data pattern $g(\varphi)$ of a model that varies as a function g dependent on the parameter φ is exactly the same as that of the model $h(\lambda) = g(f(\lambda))$. However, although the models would produce identical data with identical assumptions, model g would be considered simpler than model h because $\varphi \in$ [0, R] has a more restricted range than $\lambda \in$ [-R, R]. In sum, by a simple transformation trick any modeler could manipulate this third simplicity criterion. Note, however, that these considerations do not imply that the parameter space criterion is useless because it still does make sense to distinguish between two otherwise identical models. The rationale of this paragraph is, instead, that the criterion is useless when it is used blindly, i.e., without any mathematical consideration of the transformation of the raw parameter values in a model.

Summary of Simplicity Considerations

Although the mathematics of the NAMWR might seem fairly complex on first sight, the model adheres to the model construction principle of simplicity as indexed by the three dimensions suggested by Myung and Pitt (1997). It fits the data of Experiments 4 and 6 with far less free parameters than the NAMWR, it is very restricted in its functional form, and its parameter space is restricted to positive values for three out of four parameters. This parameter space is not transformed in any way that allows for negative functional values.

5.5.5. Falsifiability

"Never propose a theory that can be tested during your lifetime"

(astrophysicists joke mentioned by Roberts & Sternberg, 1993, p. 645).

However, if one agrees "that the limitations of given models provide the impetus for the next generation of research" (Seidenberg & Plaut, 1998, p. 236), a model without limitations that cannot be falsified is useless for future research (unless it is the true model, if one believes such a model to exist). To best provide an impetus for future research, these limitations should be made explicit and a pattern of results that falsifies the model should be defined (cf. Grainger & Jacobs, 1996).

Two limitations, however, should be distinguished. First, a model should be defined such that it can be falsified so strongly that it can be dismissed afterwards. As concerns the NAMWR, such a falsifier would be a reliable chainsaw curve form as was observed in one condition in Experiment 4 (attributed to random noise). If such a curve form were reliably observed in the absence of extreme neighborhood distributions,

the model would have to be dismissed. Second, a model might be ‘falsified’ by data that are beyond its currently limited scope. Then the model would not be falsified in general, but would need to be refined. In this way, the limitations of a model really provide an impetus for future research.

For example, the old model had no parameters for lexical processes (indexed by the frequency effect), lexical-orthographic processes (neighborhood distribution), and their interactions with word length, viewing position, and with each other. In this thesis, these limitations, as well as the lack of perceptual frequency parameter for visual word form, provided the impetus for the current experiments. The experiments falsified the old model and led to the specific incorporation of the frequency, neighborhood, word length and recognition of fixation parameters. The specific nature of this incorporation can again be tested in further experiments as well as the lack of other parameters (e.g., for phonology) in much the same way I have done in this thesis. The most important issues to test within the PIT are in my opinion i) the influence of phonology, ii) the transformation from a category-based to an item-based viewing position curve, iii) the nature of the perceptual frequency effect of visual word form, iv) the nature of the frequency interaction with other variables, and v) the more specific separation of the inhibitory influence of positional neighbors and the facilitatory influence of the number of neighbors in the current model. I will elaborate on these points in detail now.

Influence of Phonology

Virtually all major connectionist models in the 90s have made their transformation from orthographic-lexical models to orthographic-phonological-lexical models (Coltheart et al., 1993; Jacobs et al., 1998; Plaut et al., 1996; Zorzi et al., 1998). While an influence of phonology on word recognition is undisputed, it is still controversial how, at which level, and under which conditions phonology helps or hinders reading performance. For example, some models assume grapheme-phoneme regularity to be most relevant for orthography-phonology translations (Coltheart et al., 1993; Rastle & Coltheart, 1999), while others assume subsyllabic components to be most relevant (Jacobs et al., 1998; Plaut et al., 1996). Again others assume functional units that fall in neither of both categories (Richter, 1999). Experimental research suggests multiple levels to be relevant like graphemes (Rey, Ziegler, & Jacobs, 2000), subsyllabic components (onset, nucleus, and coda, Nuerk, Rey, et al., 2000; rimes or bodies ,Treiman et al., 1995; Ziegler & Perry, 1998). In future models, the level(s) at which phonology influences performance must be defined. The nature of the interaction of phonology with other variables such as viewing position and frequency must be specified. The following considerations may illustrate this issue: Phonological effects may differ in size on different viewing positions similarly to the way they do in different frequency groups.

One could assume that orthographic-phonological structure that is easy to process (i.e., a frequent and consistent /regular structure) may improve word recognition more under otherwise poor conditions than under optimal conditions. Such an interaction of phonological effects with experimental conditions has already been shown for different frequency groups – phonology effects are mostly observed for low-frequency words (e.g., Nuerk, Rey, et al., 2000). The interaction of phonology with visual determinants might show a similar effect: Possibly, phonology effects are greatest on poor viewing position. Such interactions should be systematically investigated to obtain constraints for a future incorporation for phonological properties in the NAMWR. The limitation of the model with regard to phonology may provide the impetus for future research. It may, thus, help to quantitatively specify phonological influence in a mathematical model of word recognition.

Transformation from a Category-Based to an Item-Based Approach

The old model did not distinguish between different categories of letter strings (e.g., frequent or non-frequent words). The NAMWR may be a first step toward a true word recognition model because it does just that. However, the serious shortcomings of a categorical approach have often been criticized (Balota & Spieler, 1999; Schmidt-Weigand, 1999; Spieler & Balota, 1997; Ziegler et al., 1998). In the NAMWR, this category-based approach implies two caveats. First, the model assumes that each letter at each position at each fixation distance is visually equally well recognizable. This may not be the case. For example, Paap (personal communication) pointed out that in the original version of the AVM model manuscript (Paap et al., 1982) different confusion matrices for different letter positions were used to obtain better word recognition predictions. So, the assumption that the letter legibility of a given letter can be analogously computed for each letter and viewing position may be oversimplified. Moreover, in the FRAG Experiments 5 and 7, it was demonstrated that the general confusability of the individual letters in a word played an important role. Similar effects may be observed in the PIT I used in the VPE experiments. Therefore, the confusability of each letter at each position must be investigated in a large number of trials in the very font and presentation form that is used for following word recognition experiments (I suggest at least 100 per letter per position over all participants to obtain a valid confusability matrix, cf. Schmidt-Weigandt, 1999; Ziegler et al., 1998). Only then, after a specific incorporation of individual letter confusability, are individual item predictions possible for each viewing position. However, these predictions can only be tested empirically by producing stable viewing position curves for each word by repeating its presentation over many participants (see Balota & Spieler, 1999;

Schmidt-Weigand, 1999; Spieler & Balota, 1997; Ziegler et al., 1998, for such item analyses). The limitation of the current model is clear: It only allows categorical predictions.

Perceptual Frequency Effect of Visual Word Form

As already discussed above, the NAMWR is, like any other model, not yet capable of modeling perceptual frequency effects because the perceptual learning curve and the relation of case-specific and abstract perceptual form is not yet constrained enough for formal modeling. More research is needed to find parametric solutions that can account for perceptual frequency effects.

Interaction of Frequency with Other Variables

While the above three points refer to limitations of the model that are beyond its current scope, limitations 4 and 5 question the appropriateness of specific parametric and functional assumptions of the current NAMWR. The NAMWR currently predicts stronger frequency effects for poor viewing positions thus providing successful fits for the data of Experiment 4. This prediction can be tested. Especially the assumption that frequency effects monotonously increase with deteriorating performance seems oversimplified. This assumption may no longer hold when performance becomes too poor because the alteration of the frequency effect by visual determinants may be U-shaped (i.e., a weak frequency effect under near-optimal conditions, a larger effect under poorer conditions that decreases again when conditions become too poor, see Behrman et al., 1998). Hence, further experiments with even poorer performance for low-frequency words than in the current experiment are needed to investigate this issue. Here, the postulated monotonous increase of the frequency effect is a limitation of the model and the function of its effect may need to be changed (and eventually also included in the normal distribution transformation). Note that the frequency by viewing position interaction was also not statistically significant in Experiment 6. However, the current simple linear incorporation of the lexical parameter provided successful fits for the data in all experiments including Experiment 6. Thus, this rather simple linear account of frequency effects first needs to be falsified before there is a need to change this incorporation.

Distinction between the Inhibitory Effects of Positional Neighborhood and the Facilitatory Effects of Overall Neighborhood

This distinction is a very interesting assumption which is explicitly incorporated in the NAMWR: Any single neighbor produces an inhibitory effect on the improvement of word recognition imposed by lexical processing, particularly at confusable or badly legible letter positions, while the overall number of

neighbors is assumed to be facilitatory. Neighborhood size is mostly found to be inhibitory in the PIT, but facilitatory in the NT. In the LDT, the effect is controversial (Andrews, 1997): The neighborhood size (N) effect is mostly facilitatory, but the neighborhood frequency effect is mostly inhibitory (Grainger & Jacobs, 1996; but see Andrews, 1997). In connectionist models of the MROM-family (Grainger & Jacobs, 1996; Jacobs et al., 1998; Ziegler et al., 1998), any single (active) neighbor representation can produce inhibitory activation at the lexical level, as the activated word nodes compete with and inhibit each other. However, via top-down (feed-back) connections from the word level to the letter level, neighbors also send activation to all but one letter of the target word which activate the target word again in the next cycle. Hence, in this way neighbors may also lead to higher activation in the target word via feed-back activation to the letter level. Moreover, in the LDT, early summed lexical activation (which is higher when many words are activated) determines the speed of one response process (the sigma-process). Again, more activation (possibly caused by more activated lower-frequent neighbors) leads to faster responses. Although the assumption that a single neighbor is inhibitory and the overall number of neighbors is facilitatory is first explicitly made in the NAMWR, it is already implicitly implemented in connectionist models (see Grainger & Jacobs, 1993, for early suggestions of a similar distinction).

Above I have attributed the facilitatory effect to a more familiar orthographic-phonological structure in words with more neighbors. So there is an easy way to test and falsify this assumption in the model. Words with similar frequency, same length, similar N and HFN, and similar neighborhood distribution, but different orthographic-phonological structure (e.g., different SCF or BodyN, Nuerk, Rey, et al., 2000; Ziegler & Perry, 1998) need to be presented in different stimulus groups. If the facilitatory effect of mean N would produce appropriate fits in one group, but not in another, the incorporation of mean N may need to be replaced by some parameter of orthographic-phonological structure.

All five above points may lead to falsifications of the current model as it is now that do not necessarily falsify the whole model. Instead, these possible falsifications may provide the impetus for future research and a better refined model in much the same way the Nazir et al. (1998) model provided the impetus for this thesis. However, falsifications may lead to ever new refinements of an inadequate model, so that a false and inadequate construction idea of a model may never be dropped (cf. Grainger & Jacobs, 1996; Jacobs & Grainger, 1994). Therefore, modelers need to postulate a pattern of results that falsifies the whole model family as it is. As noted above, this falsificator in the current model is its functional form. In the absence of absolutely extreme neighborhood distributions (i.e., only neighbors on the initial or the last letters in very long words), I should not observe a chainsaw distribution in any viewing position curve. In

Experiment 4, one of the curves already has a chainsaw form. I claim that this is a statistical artifact due to random noise in the data. However, if this effect was systematic, the model would already be falsified.

5.5.6. Generality

Grainger and Jacobs (1994) distinguish vertical and horizontal generality. In their definition (p. 1316) *"horizontal generality refers to a model's ability to generalize across different stimulus sets and configurations (stimulus generality), different tasks (task generality), or response types or measures (response generality)".* Vertical generality refers *"to a model's ability to generalize across different scales of the modeled process [...] or different types or sizes of the processing structure"* (pp. 1317-1318). For connectionist models, this implies that the simulation of effects in a model with relatively few nodes does still work when the number of nodes is increased. Thus, the simulations should not be specific to the selected number of nodes in the model. For parametrical models, vertical generality is harder to define. Increasing the size of a parametrical model basically implies adding additional free or fixed parameters. However, adding parameters does not simply increase the size of the processing structure. Instead, adding parameters mostly (and in the NAMWR) also implies that performance for additional stimulus configurations can be captured by the model (i.e., an overlap with horizontal generality). Moreover, increasing the size of a parametrical model also relates to the principle of nested modeling, because adding additional parameters may also imply that the old model is a special case of a new one with more parameters (see below). Because these overlaps can – in my opinion – hardly be disentangled for parametrical models, horizontal generality and the nested modeling principle will no be discussed without a separate paragraph on vertical generality.

With regard to horizontal stimulus generality, the NAMWR is more general than the old model of Nazir et al. (1998). It can fit data from any combination of letter number and fixation number, while the old model was restricted to experiments in which those two numbers were identical. Moreover, the NAMWR can account for the several effects in visual word recognition, such as word length, frequency, and neighborhood distribution effects. In particular, it can also account for the characteristic interactions of these effects with visual variables and with each other. However, as noted above, the NAMWR still has its limitations with respect to stimulus generality, e.g., with respect to the perceptual frequency of item-specific visual word form. However, this limitation does not disqualify the model because no word recognition model can do that yet as there is no adequate metric for this perceptual frequency.

One weakness of the NAMWR as well as of its direct precursor is that it possesses neither task- nor response-generality. This is an important property because Experiment 7 has shown that in a PIT (the FRAG task) one can obtain different patterns of data for different dependent variables within the same task. However, there is no mathematical model of the VPE that possesses task- or response-generality, and more generally, mathematical models are often restricted to one task (e.g., if one assumes the AVM to be a mathematical model, it is restricted to the Reicher task, Paap et al., 1982). In my view, task and response-generality is a step some connectionist modelers were able to take in the last five or ten years (Grainger & Jacobs, 1996; Jacobs et al., 1998), but which still remains a shortcoming of and a challenge for mathematical models of visual word recognition. However, if one still agrees with Estes (1975, p. 279) that *"mathematics remain our principal vehicle for the flights to imagination to smooth our experiences",* meeting this challenge may be worthwhile.

5.5.7. Nested Modeling Principle

"A model development [is needed] that is guided by the principle of nested modeling (i.e., a new model should be related to or include at least, its own, direct precursors and be tested against the old data sets [...] before testing it against new ones)" (Jacobs & Grainger, 1994, pp. 1329-1330).

Nested modeling is a principle to adhere to in model development, but not a selection criterion like the other criteria discussed above. This principle is important for model development because it prevents models growing wildly and ad hoc. Instead, models should be developed based on more reliable general core assumptions consistent within each model family.

For the most part, I included the old model as a special case of the NAMWR. Only in two cases did core assumptions of the old model have to be revised. I kept the linear drop-off of letter legibility with distance from fixation, but transformed it with the normal distribution function. This transformation was necessary because the old model could predict negative recognition values for some curves (e.g., long words), when the parameters of other curves (e.g., short words) were used. This behavior is obviously wrong when only positive accuracy values can be obtained. The new NAMWR cannot not predict negative values any more: The normal transformation now reflects the decrease in absolute legibility drop-off empirically found by Nazir et al. (1992) when recognition likelihood is close to 1 or 0. The influence of lexical knowledge (frequency and neighborhood distribution) is zero if the lexical parameter is zero. Therefore, as the old model did not assume lexical processes, it is a special case of the new model with the lexical parameter being zero. The variation of the asymmetry ratio and recognition of fixation as a

function of word length also adheres in part to the principle of nested modeling because for words with a standard length of five letters no variation of these parameters with word length is incorporated in the NAMWR. The recognition likelihood at fixation was fixed to 1 in the old model. It is now a free parameter. The reason is simple – even the letter at fixation is not recognized perfectly. Therefore, the assumption that recognition likelihood is 1 is empirically wrong. However, because there is a monotonous drop-off of letter recognition likelihood with distance from fixation, the letter on fixation is still supposed to be recognized best, but not perfectly anymore. In sum, two alterations were necessary: The normal distribution transformation to restrict the model to positive values and the assumption that the recognition at fixation parameter could vary freely. All other model refinements adhere to the principle of nested modeling, at least for the case of five-letter words.

5.5.8. Other Criteria

As introduced above a huge number of other criteria and selection principles has been introduced in the last years. Recently, Bamber and van Santen (2000) brought forward a more formalized approach in which they formally defined the following intuitive concepts: (Quantitative and qualitative) testability, identifiability, and redundancy. Their criterion testability is essentially the same as our criterion falisfiability, it the observable empirical outcome variables (in our case the viewing position curves) may produce values (i.e., empirical data patterns) that the model is unable to fit with any set of parameter values.

According to Bamber and van Santen (2000, p. 24) *"a model has redundant parameters (or redundant) observables) if, given the values of a proper subset of its parameters or observables, the values of the remaining parameters of observables can be deduced."* This definition is basically the standard definition of a linearly dependent vector in linear algebra (e.g., Fischer, 1989). There a vector $(v_1, \ldots, v_n)$ is linearly independent[21] when $a_1v_1 + \ldots + a_nv_n = 0$ only for $a_1 = \ldots = a_n = 0$ and linearly dependent when there exist other solutions. This implies that "a proper subset" of parameters can be deduced from the remaining parameters. An example for redundant observables are the probabilities of mutually exclusive and exhaustive events. The probability p_n of event n can then be computed as $p_n = 1 - (p_1 + \ldots + p_{n-1})$, i. e., as 1 minus the summed probabilities of the remaining events. Bamber and van Santen (2000, p. 25) refer to the number of observables *n* in this case as not being unique. Similar

[21] Formally, a vector family (rather than vectors or a vector) is linearly dependent or independent. The n one-dimensional vectors $((v_1, \ldots, v_n)$ constitute such a vector family. That the dimension of each of this vector is one is just the special case which Bamber and van Santen (2000) term redundancy. Colloquially, mathematicians, however, usually speak of linear dependent or independent vectors (or in this case, elements of a vector, cf. Fischer, 1989).

redundancy examples can be found for parameters. However, as Bamber and van Santen (2000, p. 25) point out, *"in practice, the potential nonuniqueness of the number of nonredundant parameters and observables does not seem to be a problem."* This is also true for the NAMWR. No parameter can simply be expressed as being a function of other parameters. The discussion of the third criterion 'identifiabilty' will also illustrate this point.

A model is identifiable "*if the values of its parameters can be ascertained from empirical observations"* (for a formal definition, see Bamber & van Santen, 2000, p. 29). Basically, this definition implies that there is only one parameter combination that is able to produce a specific output. For the NAMWR this is true, when the outcome refers to different frequency groups. Then, the lexical parameter *l* is bound by the size of the frequency effect because no other parameter has any influence on the size of this effect (for l = 0, the frequency effect is zero regardless of the values of the other parameters). The asymmetry ratio is also determined by the asymmetry of the curve: No other parameter exerts direct influence on the asymmetry of the curve. The recognition at fixation parameter r_f and the drop-off rate *d* both exert their influence on height or shape of the curve. However, this influence differs: A low rather flat curve can only be fitted by a low r_f value and a drop-off rate close to zero. In contrast, a curve that falls steeply of towards the beginning and the end of the word but has still a fairly high recognition rate on the optimal fixation position can only be fitted by a high drop-off rate, but also a fairly high recognition at fixation r_f. Thus, these qualitative consideration show that there can be empirical curves for which the parameters of the NAMWR are identifiable. In all simulations I performed, the solution was always unambiguous in that the parameter set chosen by the model was invariant from the start parameter. Nevertheless, there may be empirical curves for which the parameter solution is unidentifiable (for example, if there is no frequency effect in the data). Under which conditions the model is identifiable and under which conditions this is not the case is thus subject to a then quite complex mathematical analysis that is certainly beyond the scope of this thesis (see Bamber & van Santen, 1985, 2000, for proposals). At the current point, I suggest that formally identifiability strongly depends on the definition of and the constraints on the chosen parameter domain and, in particular, the output range (e.g., if only five viewing position points one single group of stimuli would be fitted with four parameters, the model may not be identifiable, however, if different word length, frequency, or neighborhood groups are fitted with one parameter set, I would suppose the model is identifiable). However, if the NAMWR would be successful in fitting word recognition data in future studies, at some point identifiability should be carefully studied for different output ranges.

5.6. Conclusions

In this chapter, the limitations of the old viewing position model of Nazir et al. (1998) were analyzed and a new NAMWR was suggested that overcomes or improves most of these limitations. In contrast to the old model, the NAMWR is able to successfully fit the data of different stimulus groups with an invariant set of visual parameters. It can thus be understood as a transformation from a pure VPE model to a simple word recognition model accounting for the VPE as one effect among others. This extension in scope is accompanied by a reduction in the ratio between data points and free parameters. While the old model was only able to fit five data points with two free parameters resulting in 12 parameter adjustments in Experiments 4 and 6, the NAMWR could fit 30 data points well with only four free parameters. Finally, the model refinement was discussed with respect to relevant model selection criteria and the design principle of nested modeling. Compared to the old model, the NAMWR fares better in virtually all respects of model evaluation, but is still inferior to connectionist models in some respects, such as, e.g., horizontal generality. Moreover, with two well-reasoned exceptions, the model refinement adhered to the principle of nested modeling. Therefore, the NAMWR represents a good base for parametrically studying visual word recognition. Its limitations might well provide the impetus for future experimental model tests and subsequent refinements.

6. General Summary

The guiding hypothesis of this thesis was that visual determinants of reading could be underestimated on the basis of insufficient statistical models and experimental designs. This hypothesis was confirmed in four respects: First, theoretically, reviews and critical discussion of manipulation methods and statistical models suggested that the influence of visual variables such as, e.g., item-specific word form, may not be detected when inappropriate experimental designs and analysis methods are used. Second, empirically, all three studies showed that the influence of visual variables themselves or their interaction with standard effects of visual word recognition can be detected when such appropriate designs and / or methods are used. Third, fits with the model of Nazir et al. (1998) provided additional evidence that visual parameters in that model changed in a characteristic way as a function of lexical information such as the asymmetry parameter as a function of neighborhood distribution. The specific nature of this characteristic cannot be detected by the general linear model. Fourth, the incorporation of lexical and orthographic-lexical parameters in the new NAMWR was such that these parameters interacted with visual parameters in a specified functional way. The incorporation was not simply additive and linear as suggested by the general linear model. The model fits with this functional incorporation were successful, thus indicating that

lexical effects can be well modeled when independence of visual and lexical stages or processes is no longer assumed. In sum, there is converging evidence from all four of these approaches that visual aspects of reading may be underestimated when AFL is used to interpret results in the framework of the general linear model. This shall now be elaborated for each of the studied issues.

6.1. Perceptual Learning and the Perceptual Frequency Hypothesis

"On the role of outline shape and word-specific visual patterns in identification of function words: NONE" (Title of a study of Besner, 1989)

The present study showed that the above claim of Besner and colleagues (Besner & McCann, 1987, Besner, 1989) that word-specific visual patterns do not play any role in visual word recognition does not hold. The results of this study also falsify the conclusion of Mayall et al. (1997) that reading uses visually based letter clusters of same size and same case. German nouns were better recognized when presented with initial capitalization. This effect was task-overlapping, but item-specific: It was restricted to nouns, which in German are most frequently perceived with initial capitalization. For non-nouns, no difference or the opposite effect was observed. It was suggested that the reason for the failure of earlier studies to find such effects was their choice of language (English), their visual manipulation (mostly case mixing in different frequency groups), and their inference-statistical interpretation. Null interaction effects between lexical frequency and a perceptual variable (case mixing) in an ANOVA which are interpreted using AFL are not conclusive. The present thesis demonstrates this in empirical and theoretical respect. Empirically, the finding that a German noun is recognized best in its most frequent capitalization form, i.e. with initial capitalization, contradicts conclusions and assumptions that a single (abstract) letter code is the only letter code used in visual word recognition (Besner, 1989; Mayall et al., 1997; and virtually all connectionist models). Especially instructive is Experiment 3. There, I generally found the same main perceptual frequency effect as in Experiments 1 and 2. However, I could also study the interaction between lexical frequency and perceptual frequency in Experiment 3. This interaction was not significant in any analysis. Based on such null interaction between lexical frequency and visual variables, the conclusion was often drawn that perceptual frequency did not play any role in such an experiment and in visual word recognition in general. Thus, one single experiment brought up two opposite conclusions dependent on what manipulation and what logic was used. This contradiction can only be resolved if one line of theoretical reasoning is not appropriate.

While a stable main effect between two paralleled case manipulations can hardly be neglected, the finding of a null interaction between lexical frequency and this visual manipulation can be discussed in different ways. In this thesis, it was suggested that the general linear model on which the ANOVA rests is just another (true or false) model. Other models, like interactive models, are quite successful in visual word recognition and they are also well capable of producing null interactions. A null interaction between lexical and perceptual frequency may well arise in interactive activation models when top-down influences and bottom-up influences have opposite effects which may cancel each other out. High lexical frequency may compensate for poor bottom-up information and thus produce underadditivity of lexical frequency and a perceptual manipulation. The effect of distortion of item-specific word form may be particularly large when the (perceptual) frequency of item-specific word form is high. Thus, one might observe overadditivity of (perceptual) frequency and a perceptual distortion. If lexical and perceptual frequency cannot be separated, their opposite effects might cancel each other out and additivity could be observed.

Hence, earlier conclusions that perceptual learning of item-specific visual word form does not occur probably are an artifact of applying a common, yet for this question insufficient statistical model - the general linear model. It is important to keep in mind that the inference statistics of the ANOVA are true if and only if the preconditions of the ANOVA are fulfilled. These preconditions are not only homogenous variances and normally distributed variables, but also the validness of the general linear model in general. This model assumes that psychological processes are a linear additive combination of main effects of the variables and their interactions. If null effects in an ANOVA are obtained, then this may be the case for two reasons: i) The manipulations may exert indeed no effect or ii) the general linear model may not be applicable. The second reason should not be forgotten, particularly not in fields in which alternative models, e.g., interactive activation models, were successful. Note, however, that these considerations do by no means imply that the ANOVA is an insufficient tool based on a bad model altogether. I, too, applied the ANOVA in this thesis because for inference statistics I still believe it to be the “best game in town”. However, throughout this thesis null effects and, in particular, null interaction effects were interpreted with great care and no strong theoretical conclusions were based on such null interactions. In contrast, when significant interactions or main effects were interpreted (such as a VPE or an interaction between case manipulation and syntactic class), I did not stop at suggesting that the two variables interact in some unspecific way in some black box common stage. Instead, I tried to specify the nature of the interaction qualitatively and, in the NAMWR, also quantitatively. Essentially the interaction was not considered to be the true output of a true model, but a significant interaction effect was a confirmation of a more specific

theoretical hypothesis. Altogether, the results of this thesis strongly confirm that the ANOVA should be used as an often helpful tool to test a specific hypothesis. However, one should always keep in mind that the statistical values given by an ANOVA are never meaningful by themselves, but only if the underlying statistical model is assumed to be true. This thesis demonstrates that this assumption is not always appropriate.

Besides these theoretical considerations, the content should not be forgotten. As in other domains, perceptual learning seems to be an important factor in visual word recognition in two respects. First, this study confirmed perceptual learning of location because the asymmetry of the viewing position curve for different word lengths closely corresponded to the perceptual frequency of landing positions of different word length. Second, perceptual learning of item-specific visual word form was introduced as a second source of perceptual learning in visual word recognition. Future studies should now specify the perceptual learning curve and introduce a metric of perceptual frequency that takes into account the different visual aspects of a perceived word. In sum, the two crucial assumptions of perceptual learning formulated in the introduction were confirmed: Perceptual learning takes place every time a word is encountered and it is not restricted to general perceptual units, but also item-specific.

6.2. On the Non-Additivity of Lexical and Perceptual Processes

Studying lexical frequency effects and their interactions with other variables with the principle of isolated variation does not produce conclusive results. I reviewed evidence that null interactions of frequency with visual variables and with other non-visual variables were in particular found under near-optimal presentation, reading skill or stimulus conditions. In contrast, if any of these conditions is non-optimal, an interaction of frequency with other variables seemed to become more likely. This hypothesis was also confirmed in my thesis. On central (near-optimal) viewing position, no significant interaction of frequency with word length was observed, while on edge viewing positions such an interaction was observed. Moreover, the frequency effect in Experiment 4 was larger on inferior viewing positions. Experiment 5 validated the assumption that frequency can interact with visual variables in appropriate experimental designs in a fragmentation task in which frequency and visual letter confusability are known to be the greatest sources of variance (Ziegler et al., 1998). Again, frequency exerted its largest effects in poor visual conditions. Theoretically, the functional architecture of the NAMWR supported the claim that lexical processes become more important when visual salience becomes poorer. The lexical parameter is incorporated such that frequency helps more on poorer viewing positions than on near-optimal positions. Given the specific nature of the interaction of frequency with viewing position, an additive linear

parameter would not have worked. The incorporation provided successful fits for the data, thus supporting a non-additive interpretation of the frequency effect. However, this thesis also showed that the interaction of frequency with visual variables may not always be picked up. There was no significant interaction of viewing position with frequency in Experiment 6, in which the variance in performance between different visual conditions was not as large as in Experiment 4. Nevertheless, the NAMWR with a non-additive incorporation of a lexical parameter fitted the data of Experiment 6 successfully. Again, this apparent contradiction supports the view that null interactions of frequency with other variables in an ANOVA must be handled with great care. In sum, the results of this thesis suggest that the interaction of frequency with other variables cannot be studied with isolated variation and then be generalized on the reading process. This is especially the case when participant, stimulus and presentation conditions are near-optimal, e.g., in laboratory experiments with student participants with unlimited central presentation of a single word. In a more general way, these findings call for an approach to visual word recognition in which visual variables such as viewing position or letter legibility are always manipulated rather than held constant at a high salience level.

6.3. On the Non-Additivity of Orthographic-Lexical and Perceptual Processes

The results of this thesis support the view that even the orthographic relations between lexical entries cannot be investigated without taking their visual-orthographic relations into account. Study 3 suggested that the neighborhood size effect is not a purely orthographic-lexical effect, but rather a visual-orthographic-lexical effect. Again, I suggested that when studies failed to find certain orthographic effects (such as a null effect of neighborhood position), it was due to the inappropriate application of the principle of isolated variation. Experiment 6 showed that the effects of neighborhood distribution systematically changed as a function of fixation position. Again, for central fixation position (i.e., principle of isolated variation) different neighborhood distributions did not differ as much as for other fixation positions.

A final word is necessary on the criticized principle of isolated variation whose application may mislead interpretations. This principle has brought psychology and other sciences forward for more than 100 years, and, obviously it cannot be the intention of a PhD thesis to claim that the principle itself should be dropped. Such a claim has never been made in this thesis. However, I argued and presented evidence that the application and interpretation of the principle of isolated variation was not always appropriate in a specific field of visual word recognition research. That such an application may not always appropriate is by new means a new or revolutionary thought. This assumption is implicitly already made in every 2 $\times$ 2 design in which researchers look for an interaction in an ANOVA. Such an interaction between two factors

cannot be discovered when only one of these factors is manipulated. Basically, this thesis investigated such interactions of lexical or orthographic-lexical factors with visual factors. My criticism brought forward in this thesis was, that such interactions of visual factors with others, in particular, are often not considered: The assumption is made that higher cognitive (orthographic-lexical or lexical) processes can be studied under one certain visual condition and that the obtained effects with the isolated variable (e.g., frequency or neighborhood distribution) can be simply generalized to the reading process (i.e., to any visual condition). This assumption is not compatible with the results of this thesis.

However, one could argue that already any 2 $\times$ 2 design as well as any more complex categorical design represents an extension of the principle of isolated variation. Instead of one variable, two or more variables are then “isolated” and manipulated between certain levels (in my thesis not only two, but in each study also sometimes three levels, e.g. word forms in Study 1, word lengths in Study 2, and frequency groups in Study 3). However, many variables not manipulated are held constant at a certain level (see all stimulus tables for examples of variables held constant). Thus, it is not the principle itself that one or more variables are manipulated while others are held constant that is criticized. Rather, it is the appropriateness of the application of this principle and, in particular, the generalizeability of the results that is questioned. In my thesis, I tried to be careful with respect to generalizations of the results. In all three studies, I tried to cross-validate the effect found under one specific presentation condition or manipulation to other presentation conditions or manipulations to extend the generalizeability of the results.

In such a cross-validation, Experiment 7 validated and extended the account of neighborhood being a similarity index in visual-orthographic-lexical space. Neighbors with highly confusable letter exchanges consistently produced inhibitory effects compared to hermits in all analyses and all dependent variables. In contrast, neighbors with lowly confusable letter exchanges did not produce such consistent effects; sometimes descriptively even facilitatory effects (compared to hermits) were observed. In direct comparison, words with highly confusable neighbors produced inferior performance than words with low confusable neighbors.

There is a controversial discussion about which neighbors are facilitatory or inhibitory in which task (Andrews, 1997; Grainger & Jacobs, 1996; Ziegler & Perry, 1998). My results suggest that facilitatory effects of neighbors found in some studies are an artifact of near optimal presentation and participant skill conditions. Put more specifically, neighborhood size may only be facilitatory in the NT and the LDT because mostly near-optimal presentation conditions (e.g., central fixation position) are used in highly

skilled participants. I hypothesize that if initial presentation occurred on inferior viewing positions, the pattern of results would change. Recent studies examining neighborhood size effects in the LDT and in natural reading provided results consistent with that hypothesis (Pollatsek et al., 1999).

This hypothesis is also supported by the successful incorporation of neighborhood distribution in the NAMWR. A single positional neighbor tends to be more inhibitory when visual salience is poorer. In contrast, the overall number of neighbors is facilitatory. This facilitatory influence in a PIT might be due to more frequent sublexical phonographic units that are associated with a larger number of neighbors (Nuerk, Rey, et al., 2000; Ziegler & Perry, 1998). This functional assumption corresponds to the architecture of some interactive activation models (e.g., Jacobs et al., 1998) in which each neighbor unit does inhibit the target word representation via lateral inhibition, but may also facilitate the target word unit via excitatory top-down connections. However, my study showed that the inhibitory influence of each neighbor representation is reduced as the visual legibility of the critical letters improves. Thus, studying neighborhood effects under near-optimal presentation conditions systematically reduces inhibitory influences of neighbors to a minimum. With inhibition reduced to a minimum, the facilitatory attributes of neighbors become relatively more important and may then dominate the effects observed in empirical research.

Consequently, these results suggest that neighborhood effects and their interaction with other variables cannot be generally studied with isolated variation and then be generalized to the reading process. In particular, the common method of presenting a single word on central (near-optimal) fixation position in skilled participants is insufficient because those very conditions may most favor facilitatory effects of neighborhood. The third study calls again for an approach to visual word recognition in which visual variables are always manipulated rather than only held constant.

6.4. Final Conclusions

The conclusions for future research are thus straightforward: Visual aspects of visual word recognition should not be kept constant at a high salience level when other effects are studied because these other effects or their interaction with visual variables may be underestimated: Null effects may be observed that are specific to optimal presentation conditions. Thus, word recognition models that try to simulate higher lexical, or sublexical, orthographic or phonological effects in visual word recognition do not provide general accounts of these effects if these models have no incorporation of visual attributes. However, following Seidenberg and Plaut (1998), such limitations in those models may provide an impetus for future research. The main impetus for this thesis came from the model of Nazir et al. (1998), which was

able to fit an important visual effect to empirical data, but could not yet fit lexical and orthographic effects that were beyond its scope. The NAMWR is a first mathematical fitting model that tries to reconcile both the adequate description and visual effects and that of lexical, orthographic-lexical and word length effects in a given task. Most importantly, it tries to specify the nature of the interaction of these effects. Thus, it represents an alternative theoretical mathematical framework to that of the general linear model which is, however, restricted to the small domain of word recognition in a PIT and does not allow inference-statistical conclusions yet. Nevertheless, its good fitting performance strengthens the claim that the general linear model may not necessarily be the true or most appropriate mathematical model in this domain.

Word recognition researchers have been quite successful in the last 30 years in investigating and understanding higher lexical and sublexical, orthographic and phonological processes under particular presentation conditions. However, I believe that it is time to relate this research to the investigation of the basic visual attributes of visual word recognition because a normal reading process does not always occur under near-optimal visual presentation of a single word to a skilled reader. The NAMWR is a model that tries to specify this relation for a PIT. It may hopefully be an impetus for a next generation of research reconciling basic visual and higher cognitive processes in visual word recognition to gain a better insight on the normal reading process.

VI. References

Aghababian, V., & Nazir, T. A. (2000). Developing normal reading skills: Aspects of visual processes underlying word recognition. *Journal of Experimental Child Psychology, 76,* 123-150.

Ahissar, M., & Hochstein, S. (1993). Attentional control of early perceptual learning. *Proceedings of the National Academy of Sciences of the USA, 90,* 5718-5722.

Ahissar, M., & Hochstein, S. (1996). Learning pop-out detection: Specificities to stimulus characteristics. *Vision Research, 36,* 3487-3500.

Ahissar, M., & Hochstein, S. (1997). Task difficulty and the specificity of perceptual learning. *Nature, 387,* 401-406.

Ahissar, M., Laiwand, R., Kozminsky, G., & Hochstein, S. (1998). Learning pop-out detection: Building representations for conflicting target-distractor relationships. *Vision Research, 38,* 3095-3107.

Allen, P. A., Wallace, B., & Weber, T. A. (1995). Influence of case type, word frequency, and exposure duration on visual word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 21,* 914-934.

Anderson, N. H. (1981). *Foundations in information integration theory.* New York: Academic Press.

Andrews, S. (1989). Frequency and Neighborhood Effects on Lexical Access: Activation or Search? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 802-814.

Andrews, S. (1996). Lexical retrieval and selection processes: Effects of transposed-letter confusability. *Journal of Memory and Language, 35,* 775-800.

Andrews, S. (1997). The effect of orthographic similarity on lexical retrieval: Resolving neighborhood conflicts. *Psychonomic Bulletin and Review, 4,* 439-461.

Arguin, M., & Bub, D. (1995). Priming and response selection processes in letter classification and identification tasks. *Journal of Experimental Psychology: Human Perception and Performance, 21,* 1199-1219.

Baayen, R. H., Piepenbrock, R., & van Rijn, H. (1993). *The CELEX Lexical Database (CD-ROM).* Philadelphia, PA, Lingusitic Data Consortium, University of Pennsylvania.

Backman, J., Bruck, M., Hebert, M., & Seidenberg, M. S. (1984). Acquisition and use of spelling-sound correspondences in reading. *Journal of Experimental Child Psychology, 38,* 114-133.

Balota, D. A., & Chumbley, J. I. (1984). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human Perception and Performance, 10,* 340-357.

Balota, D. A., & Spieler, D. H. (1999). Word frequency, repetition, and lexicality effects in word recognition tasks: Beyond measures of central tendency. *Journal of Experimental Psychology: General, 128,* 32-55.

Bamber, D., & van Santen, J. P. H. (1985). How many parameters can a model have and still be testable. *Journal of Mathematical Psychology, 29,* 443-473.

Bamber, D., & van Santen, J. P. H. (2000). How to assess a model's testability and identifiability. *Journal of Mathematical Psychology, 44,* 20-40.

Banich, M. T. (1997). *Neuropsychology: The neural bases of mental function.* Boston, MA: Houghton Mifflin.

Beauvillain, C., Doré, K., & Baudouin, V. (1996). The "center of gravity" of words: Evidence for an effect of the word-initial letters. *Vision Research, 36,* 589-603.

Becker, C. A. (1979). Semantic context and word frequency effects in visual word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 5,* 252-259.

Becker, C. A., & Killion, T. H. (1977). Interaction of visual and cognitive effects in word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 3,* 389-401.

Behrmann, M., Plaut, D. C., & Nelson, J. A. (1998). A literature review and new data supporting an interactive account of letter by letter reading. *Cognitive Neuropsychology, 15,* 7-51.

Besner, D. (1989). On the role of outline shape and word-specific visual pattern in the identification of function words: NONE. *Quarterly Journal of Experimental Psychology, 41A,* 91-105.

Besner, D., Davelaar, E., Alcott, D., & Parry, P. (1984). Wholistic reading of alphabetic print: Evidence from FBM and FBI. In L. Henderson (Ed.), *Orthographies and reading* (pp. 121-135). Hillsdale, NJ: Erlbaum.

Besner, D., & Johnston, J. C. (1989). Reading and the mental lexicon: On the interaction of visual, orthographic, phonological, and lexical information. In W. Marslen-Wilson (Ed.), *Lexical processes and representation* (pp. 291-316). Cambridge, MA: MIT Press.

Besner, D., & McCann, R. S. (1987). Word frequency and pattern distortion in visual word identification and producion: An examination of four classes of models. In Coltheart (Ed.), *Attention and Performance XII: The psychology of reading* (pp. 201-219). London: Academic Press.

Besner, D., & Smith, M. C. (1992). Models of visual word recognition: When obscuring the stimulus yields a clearer view. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18* , 468-482.

Borowsky, R., & Besner, D. (1993). Visual word recognition: A multistage activation model. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 813-840.

Bosman, A. M. T., & Van Orden, G. C. (1997). Why spelling is more difficult than reading. In C. -A. Perfetti, L. Rieben, & M. Fayol (Eds.), *Learning to spell: Research, theory, and practice across languages* (pp. 173-194). Mahwah, NJ, USA: Lawrence Erlbaum Associates, Inc., Publishers.

Bouma, H., & Legein, C. P. (1977). Foveal and parafoveal recognition of letters and words by dyslexics and by average readers. *Neuropsychologia, 15,* 69-80.

Bouwhuis, D., & Bouma, H. (1979). Visual word recognition of three-letter words as derived from the recognition of the constitutent letters. *Perception and Psychophysics, 25,* 12-22.

Bradshaw, J. L., & Nettleton, N. C. (1983). *Human Cerebral Asymmetry.* Englewood Cliffs, NJ: Prentice-Hall .

Brown, J. S., & Carr, T. H. (1993). Limits on perceptual abstraction in reading: Asymmetric transfer between surface forms differing in typicality. *Journal of Experimental Psycholgy: Learning, Memory, and Cognition, 19,* 1277-1296.

Brysbaert, M., Vitu, F., & Schroyens, W. (1996). The right visual field advantage and the optimal viewing position effect: On the relation between foveal and parafoveal word recognition. *Neuropsychology, 10,* 385-395.

Butler, B., & Hains, S. (1979). Individual differences in word recognition latency. *Memory and Cognition, 7,* 68-76.

Carreiras, M., Perea, M., & Grainger, J. (1997). Effects of orthographic neighborhood in visual word recognition: Cross-task comparisons. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 857-871.

Chambers, S. M. (1979). Letter and order information in lexical access. *Journal of Verbal Learning and Verbal Behavior, 18,* 225-241.

Collyer, C. E. (1985). Goodness-of-fit patterns in a computer cross-validation procedure comparing a linear and a threshold model. *Behavior Research Methods, Instruments, and Computers, 18,* 618-622.

Colombo, L. (1986). Activation and inhibition with orthographically similar words. *Journal of Experimental Psychology: Human Perception and Performance, 12,* 226-234.

Coltheart, M., Curtis, B., Atkins, P., & Haller, H. (1993). Dual-route and parallel-distributed processing approaches. *Psychological Review, 100,* 589-608.

Coltheart, M., Davelaar, E., Jonasson, J. T., & Besner, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), *Attention and performance VI* (pp. 151-216). London: Academic Press.

Coltheart, M., & Rastle, K. (1994). Serial processing in reading aloud: Evidence for dual-route models of reading. *Journal of Experimental Psychology: Human Perception and Performance, 20*, 1197-1211.

Corcoran, D. W. J. (1966). An acoustic factor in letter cancelation. *Nature, 210,* 658.

Cutting, J. E., Bruno, N., Brady, N. P., & Moore, C. (1992). Selectivity, scope, and simplicity of models: A lesson from fitting judgments of perceived depth. *Journal of Experimental Psychology: General, 121,* 364-381.

De Luca, E., & Fahle, M. (1999). Learning of interpolation in 2 and 3 dimensions. *Vision Research, 39,* 2051-2062.

Deubel, H., O'Regan, J. K., & Radach, R. (2000). Attention, information processing, and eye movement control. In A. Kennedy, R. Radach, D. Heller, & J. Pynte (Eds.), *Reading as a perceptual process* (355-374). Oxford: Elsevier.

Deutsch, A., & Rayner, K. (1999). Initial fixation position in reading Hebrew words. *Language and Cognitive Processes, 14,* 393-421.

Drewnowski, A., & Healy, A. F. (1977). Detection errors on *the* and *and*: Evidence for reading unit larger than the word. *Memory and Cognition, 5,* 636-647.

Drewnowski, A., & Healy, A. F. (1980). Missing *-ing* in reading: Letter detection errors on word endings. *Journal of Verbal Learning and Verbal Behavior, 19,* 247-262.

Estes, W. K. (1975). Some targets for mathematical psychology. *Journal of Mathematical Psychology, 12,* 263-282.

Evett, L. J., & Humphreys, G. W. (1981). The use of abstract graphemic information in lexical access. *Quarterly Journal of Experimental Psychology, 33A,* 325-350.

Fahle, M., & De Luca, E. (1994). Spatio-temporal interpolation in depth. *Vision Research, 34,* 343-348.

Fahle, M., & Edelman, S. (1993). Long-term learning in vernier acuity: Effects of stimulus orientation, range and of feedback. *Vision Research, 33,* 397-412.

Farid, M., & Grainger, J. (1996). How initial fixation position influences visual word recognition: A comparison of French and Arabic. *Brain and Language, 53,* 351-368.

Fischer, G. (1989). *Lineare Algebra* (9th Ed.). Braunschweig, Germany: Vieweg.

Forster, K. I. (1990). Lexical processing. In D. N. Osherson & H. Lasnitz (Eds.), *Language: An invititation to cognitive science* (pp. 95-133). Cambridge, MA.

Forster, K. I., & Chambers, S. M. (1973). Lexical access and naming time. *Journal of Verbal Learning and Verbal Behaviour, 12,* 627-635.

Forster, K. I., & Davis, C. (1991). The density constraint on form-priming in the naming task: Interference effects from a masked prime. *Journal of Memory and Language, 30,* 1-25.

Forster, K. I., & Shen, D. (1996). No enemies in the neighborhood: Absence of inhibitory neighborhood effects in lexical decision and semantic categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 696-713.

Forster, K. I., & Taft, M. (1994). Bodies, antibodies, and neighborhood-density effects in masked form priming. *Journal of Experimental Psychology: Learning: Memory, and Cognition, 20,* 844-863.

Forster, M. R. (2000). Key concepts in model selection: Performance and generalizability. *Journal of Mathematical Psychology, 24*, 205-231.

Frauenfelder, U. H., Baayen, R. H., & Hellwig, F. M. (1993). Neighborhood density and frequency across languages and modalities. *Journal of Memory and Language, 32,* 781-804.

Frederiksen, J. R. (1978). Assessment of lexical decoding and lexical skills in their relation to reading proficiency. In A. M. Lesgold, J. W. Pellegrino, S. D. Fokkema, & R. Glaser (Eds.), *Cognitive psychology and instruction* (pp. 153-169). New York: Plenum Press.

Frederiksen, J. R., & Kroll, J. F. (1976). Spelling and sound: Approaches to the internal lexicon. *Journal of Experimental Psychology: Human Perception and Performance, 2,* 361-379.

Gibbs, P., & Van Orden, G. C. (1998). Pathway selection's utility for control of word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 1162-1187.

Glushko, R. J. (1979). The organization and activation of orthographic knowledge in reading aloud. *Journal of Experimental Psychology: Human Perception and Performance, 5,* 674-691.

Graf, R., & Nuerk, H.-C. (1999). Phonographische Nachbarn in der Benennungsaufgabe [Phonographic neighbors in the naming task]. In E. Schröger, A. Mecklinger, & A. Widmann (Eds.), *Experimentelle Psychologie: Abstracts der 41. Tagung experimentell arbeitender Psychologen* (p. 352). Lengerich: Pabst Science Publishers.

Grainger, J. (1990). Word frequency and neighborhood frequency effects in lexical decision and naming. *Journal of Memory and Language, 29,* 228-244.

Grainger, J., & Jacobs, A. M. (1993). Masked partial-word priming in visual word recognition: Effects of positional letter frequency. *Journal of Experimental Psychology: Human Perception and Performance, 19,* 951-964.

Grainger, J., & Jacobs, A. M. (1994). A dual read-out model of word context effects in letter perception: Further investigations of the word superiority effect. *Journal of Experimental Psychology: Human Perception and Performance, 20,* 1158-1176.

Grainger, J., & Jacobs, A. M. (1996). Orthographic processing in visual word recognition: A multiple read-out model. *Psychological Review, 103,* 518-565.

Grainger, J., O'Regan, J. K., Jacobs, A. M., & Segui, J. (1989). On the role of competing word units in visual word recognition: The neighborhood frequency effect. *Perception and Psychophysics, 45,* 189-95.

Grainger, J., O'Regan, J. K., Jacobs, A. M., & Segui, J. (1992). Neighborhood frequency effects and letter visibility in visual word recognition. *Perception and Psychophysics, 51,* 49-56.

Greenberg, S. N., & Koriat, A. (1991). The missing-letter effect for common function words depends on their linguistic function in the phrase. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 1051-1061.

Greenberg, S. N., Koriat, A., & Shapiro, A. (1992). The effects of syntactic structure on letter detection in adjacent function words. *Memory and Cognition, 20,* 663-70.

Haber, L. R., Haber, R. N., & Furlin, K. R. (1983). Word length and word shape as sources of information in reading. *Reading Research Quarterly, 18,* 165-189.

Haber, R. N., & Haber, L. R. (1981). The shape of a word can specify its meaning. *Reading Research Quarterly, 16,* 334-345.

Hadley, J. A., & Healy, A. F. (1991). When are reading units larger than the letter? Refinement of the unitization reading model. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 1062-1073.

Healy, A. F. (1976). Detection errors on the word the: Evidence for reading units larger than letters. *Journal of Experimental Psychology: Human Perception and Performance, 2,* 235-42.

Healy, A. F. (1980). Proofreading errors in the word *The*: New evidence on reading units. *Journal of Experimental Psychology: Human Perception and Performance, 6,* 45-57.

Healy, A. F., Conboy, G. L., & Drewnowski, A. (1987). Characterizing the processing units of reading. In B. K. Brittion & S. M. Glynn (Eds.), *Executive control processes in reading* (pp. 279-296). Hillsdale, NJ: Erlbaum.

Healy, A. F., & Drewnowski, A. (1983). Investigating the boundaries of reading units: Letter detection in misspelled words. *Journal of Experimental Psychology: Human Perception and Performance, 9,* 413-426.

Henderson, L., & Chard, M. J. (1976). On the nature of the facilitation of visual comparisons by lexical membership. *Bulletin of the Psychonomic Society, 7,* 432-434.

Hudson, P. T., & Bergman, M. W. (1985). Lexical knowledge in word recognition: Word length and word frequency in naming and lexical decision tasks. *Journal of Memory and Language, 24,* 46-58.

Huntsman, L. A., & Lima, S. D. (1996). Orthograhic neighborhood structure and lexical access. *Journal of Psycholinguistic Research, 25,* 417-429.

Hyönä, J. (1995). An eye movement analysis of topic-shift effects dureing repeated reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 1365-1373.

Hyönä, J., & Pollatsek, A. (1998). Reading Finnish compound words: Eye fixations are affected by component morphemes. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 1612-1627.

Iacoboni, M., & Zaidel, E. (1996). Hemispheric independence in word recognition: Evidence from unilateral and bilateral presentations. *Brain and Language, 53,* 121-140.

Inhoff, A. W., & Rayner, K. (1986). Parafoveal word processing during eye fixations in reading: Effects of word frequency. *Perception and Psychophysics, 40,* 431-439.

Inhoff, A. W., & Tousman, S. (1990). Lexical priming from partial-word previews. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 825-836.

Jacobs, A. M., & Grainger, J. (1991). Automatic letter priming in an alphabetic decision task. *Perception and Psychophysics, 49,* 43-52.

Jacobs, A. M., & Grainger, J. (1992). Testing a semistochastic variant of the interactive activation model in different word recognition experiments. *Journal of Experimental Psychology: Human Perception and Performance, 18,* 1174-1188.

Jacobs, A. M., & Grainger, J. (1994). Models of visual word recognition: Sampling the state of the art. *Journal of Experimental Psychology: Human Perception and Performance, 20,* 1311-1334.

Jacobs, A. M., Grainger, J., Rey, A., & Ziegler, J. C. (1998). MROM-P: An interactive activation, multiple read-out model of orthographic and phonological processes in visual word recognition. In J. Grainger & A. M. Jacobs (Eds.), *Localist connectionist approaches to human cognition* (pp. 147-188). Hillsdale, NJ: Lawrence Erlbaum Associates.

Jared, D. (1997). Spelling-sound consistency affects the naming of high-frequency words. *Journal of Memory and Language, 36,* 505-529.

Jared, D., McRae, K., & Seidenberg, M. S. (1990). The basis of consistency effects in word naming. *Journal of Memory and Language, 29,* 687-715.

Jared, D., & Seidenberg, M. S. (1991). Does word identification proceed from spelling to sound to meaning? *Journal of Experimental Psychology: General, 120,* 358-394.

Jesse, A., Nuerk, H.-C., Graf, R., & Massaro, D. (1999, April). *Konsistenzeffekte im Deutschen [Consistency effects in German].* Poster presented on the 41. Tagung experimentell arbeitender Psychologen. Leipzig, Germany.

Karni, A., & Sagi, D. (1991). Where practice makes perfect: Evidence for primary visual cortex plasticity. *Proceedings of the National Academy of Science, 88,* 4966-4970.

Kinoshita, S. (1987). Case alternation effect: Two types of word recognition? *Quarterly Journal of Experimental Psychology, 39A,* 701-720.

Koriat, A., & Greenberg, S. N. (1991). Syntactic control of letter detection: Evidence from English and Hebrew nonwords. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17* , 1035-1050.

Koriat, A., Greenberg, S. N., & Goldshmid, Y. (1991). The missing-letter effect in Hebrew: Word frequency or word function? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 66-80.

Koriat, A., & Norman, J. (1985). Reading rotated words. *Journal of Experimental Psychology: Human Perception and Performance, 11,* 490-508.

Landerl, K., Wimmer, H., & Frith, U. (1997). The impact of orthographic consistency on dyslexia: A German-English comparison. *Cognition, 63,* 315-34.

Lee, C. H. (1999). A locus of the word-length effect on word recognition. *Reading Psychology, 20,* 129-150.

Loftus, G. R., & Masson, M. (1994). Using confidence intervals in within subject designs. *Psychonomic Bulletin and Review, 1,* 476-490.

Logan, G. D., & Cowan, W. B. (1984). On the ability to inhibit thought and action. *Psychological Review, 91,* 295-327.

Lorch, R. F. (1986). Use of a word reading task for studying word recognition. *Bulletin of the Psychonomic Society, 24,* 11-14.

Massaro, D. W., & Cohen, M. M. (1993). The paradigm and the fuzzy logical model of perception are alive and well. *Journal of Experimental Psychology: General, 122,* 115-124.

Massaro, D. W., & Cohen, M. M. (1994). Visual, orthographic, phonological, and lexical influences in reading. *Journal of Experimental Psychology: Human Perception and Performance, 20,* 1107-1128.

Massaro, D. W., Jastrzembski, J. E., & Lucas, P. A. (1981). Frequency, orthographic regularity, and lexical status in letter and word perception. *The Psychology of Learning and Motivation, 5,* 163-200.

Massaro, D. W., Taylor, G. A., Venezky, R. L., Jastrzembki, J. E., & Lucas, P. A. (1980). *Letter and word perception: Orthograhic structure and visual processing in reading .* New York: North-Holland.

Mayall, K., & Humphreys, G. W. (1996a). A connectionist model of alexia: Covert recognition and case mixing effects. *British Journal of Psychology, 87,* 355-402.

Mayall, K., & Humphreys, G. W. (1996b). Case mixing and the task sensitive disruption of lexical processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 278-294.

Mayall, K., Humphreys, G. W., & Olson, A. (1997). Disruption to word or letter processing? The origins of case-mixing effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 1275-86.

McClelland, J. L. (1979). On the time relations of mental processes: An examination of systems of processes in cascade. *Psychological Review, 86,* 287-330.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: I. An account of basic findings. *Psychological Review, 88,* 375-407.

McConkie, G. W., Kerr, P. W., Reddix, M. D., Zola, D., & Jacobs, A. M. (1989). Eye movement control during reading: II. Frequency of refixating a word. *Perception and Psychophysics, 46,* 245-53.

McConkie, G. W., & Zola, D. (1979). Is visual information integrated accross successive integrations of reading? *Perception and Psychophysics, 25,* 221-224.

Montant, M., Nazir, T. A., & Poncet, M. (1998). Pure alexia and the viewing position effect in printed words. *Cognitive Neuropsychology, 15,* 93-140.

Myung, I. J. (2000). The importance of complexity in model selection. *Journal of Mathematical Psychology, 44,* 190-204.

Myung, I. J., Forster, M. R., & Browne, M. W. (2000). Guest editors‘ introduction: Special issue on model selection. *Journal of Mathematical Psychology, 44,* 1-2.

Myung, I. J., & Pitt, M. A. (1997). Applying Occam's razor in modeling cognition: A Bayesian approach. *Psychonomic Bulletin and Review, 4,* 79-95.

Nazir , T. A. (1991). On the role of refixations in letter strings: The influence of oculomotor factors. *Perception and Psychophysics, 49,* 373-389.

Nazir, T. A. (2000). Traces of print along the visual pathway. In A. Kennedy, R. Radach, D. Heller, & J. Pynte (Eds.), *Reading as a perceptual process* (pp. 3-22). Oxford: Elsevier.

Nazir, T. A., Deutsch, A., Grainger, J., & Frost, R. (submitted). *Perceptual learning in reading: Evidence from readers of Roman and Hebrew scripts.* Manuscript submitted for publication.

Nazir, T. A., Heller, D., & Sussmann, C. (1992). Letter visibility and word recognition: The optimal viewing position in printed words. *Perception and Psychophysics, 52,* 315-28.

Nazir, T. A., Jacobs, A. M., & O'Regan, J. K. (1998). Letter legibility and visual word recognition. *Memory and Cognition, 26,* 810-821.

Nazir, T. A., O'Regan, J. K., & Jacobs, A. M. (1991). On words and their letters. *Bulletin of the Psychonomic Society, 29,* 171-174.

Nißlein, M., Müsseler, J., & Koriat, A. (1999). Inkongruente Groß- und Kleinschreibung und der Missing Letter Effect [Incongruent capitalization and decapitalization and the missing letter effect]. In E. Schröger, A. Mecklinger, & A. Widmann (Eds.), *Experimentelle Psychologie: Beiträge zur 41. Tagung experimentell arbeitender Psychologen* (p. 139). Lengerich: Pabst Science Publishers.

Norris, D. (1984). The effects of frequency, repetition and stimulus quality in visual word recognition. *Quarterly Journal of Experimental Psychology, 36A,* 507-518.

Nuerk, H.-C., Graf, R., Boecker, M., Gauggel, S., & Jacobs, A. M. (in press). Der Wortfragmentationstest FRAG als computergestütztes Verfahren zur Erfassung der visuellen Worterkennung bei aphasischen und nicht-aphasischen Patienten [The word fragmentation test as a computerized test to assess visual word recognition in aphasic and non-aphasic patients]. *Zeitschrift für Neuropsychologie.*

Nuerk, H.-C., John, J., Jacobs A. M, Schulte-Koerne, G., Graf, R., & Gauggel, S. (2000). Lexikalische und sublexikalische Prozesse beim Auftreten des Blickpositionseffekts in der normalen Leseentwicklung und bei Legasthenikern [Lexical and sublexical processes determining the viewing position effect in normal and dyslectic readers]. In D. Vorberg, A. Fuchs, T. Futterer, A. Heinecke, U.

Heinrich, U. Mattler, & S. Töllner (Eds.), *Experimentelle Psychologie: 42. Tagung experimentell arbeitender Psychologen* (p. 68). Lengerich: Pabst Science Publishers.

Nuerk, H.-C., Rey, A., Graf, A., & Jacobs, A. M. (2000). Phonographic units in visual word recognition. *Current Psychology Letters: Behavior, Brain and Cognition, 2,* 25-36.

O'Regan, J. K., Levy-Schoen, A., Pynte, J., & Brugaillere, B. (1984). Convenient fixation location within isolated words of different length and structure. *Journal of Experimental Psychology: Human Perception and Performance, 10,* 250-257.

O'Regan, J. K., & Jacobs A. M. (1992). Optimal viewing position effect in word recognition: A challenge to current theory. *Journal of Experimental Psychology: Human Perception and Performance, 18,* 185-197.

Oden, G. C., & Massaro, D. W. (1978). Integration of featural information in speech perception. *Psychological Review, 85,* 172-191.

Olzak, L. A., & Thomas, J. P. (1986). Seeing spatial patterns. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance*. New York: Wiley.

Osterhout, L., Bersick, M., & McKinnon, R. (1997). Brain potentials elicited by words: Word length and frequency predict the latency of an early negativity. *Biological Psychology, 46,* 143-168.

Oßwald, K., Brustmann, D., Nazir, T. A., Nuerk, H.-C., Graf, R., & Jacobs A. M. (1998). Die optimale Blickposition bei lese-rechtschreibschwachen Kindern [Optimal viewing position in dyslexic children]. In H. Lachnit, A. M. Jacobs, & F. Rösler (Eds.), *Experimentelle Psychologie: Abstracts der 40. Tagung experimentell arbeitender Psychologen* (pp. 252). Lengerich: Pabst Science Publishers.

Paap, K. R., & Johansen, L. (1994). The case of the vanishing frequency effect: A retest of the verification model. *Journal of Experimental Psychology: Human Perception and Performance, 20,* 1129-1157.

Paap, K. R., & Newsome, S. L. (1980). A perceptual-confusion account of the WSE in the target search paradigm. *Perception and Psychophysics, 27,* 444-456.

Paap, K. R., Newsome, S. L., McDonald, J. E., & Schvaneveldt, R. W. (1982). An activation--verification model for letter and word recognition: The word-superiority effect. *Psychological Review, 89,* 573-94.

Paap, K. R., Newsome, S. L., & Noel, R. W. (1984). Word shape's in poor shape for the race to the lexicon. *Journal of Experimental Psychology: Human Perception and Performance, 10,* 413-28.

Peereman, R., & Content, A. (1997). Orthographic and phonological neighborhoods in naming: Not all neighbors are equally influential in orthographic space. *Journal of Memory and Language, 37,* 382-410.

Perea, M., & Pollatsek, A. (1998). The effects of neighborhood frequency in reading and lexical decision. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 767-779.

Perfetti, C. A., Goldman, S. R., & Hogaboam, T. W. (1979). Reading skill and the identification of words in discourse context. *Memory and Cognition, 7,* 273-282.

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review, 103,* 56-115.

Plourde, C. E., & Besner, D. (1997). On the locus of the word frequency effect in visual word recognition. *Canadian Journal of Experimental Psychology, 51,* 181-194.

Pollatsek, A., Perea, M., & Binder, K. S. (1999). The effects of „neighborhood size“ in reading and lexical decision. *Journal of Experimental Psychology: Human Perception and Performance, 25,* 1142-1158.

Popper, K. R. (1935). *Logik der Forschung.* Berlin: Springer.

Pynte, J. (1996). Lexical control of within-word eye movements. *Journal of Experimental Psychology: Human Perception and Performance, 22,* 958-969.

Rastatter, M. P., & Loren, C. (1988). Vocal reaction times to tachistoscopically presented high- and low-frequency verbs: Some evidence for selective minor hemisphere linguistic analysis. *Perceptual and Motor Skills, 66,* 803-810.

Rastle, K., & Coltheart, M. (1999). Serial and strategic effects in reading aloud. *Journal of Experimental Psychology: Human Perception and Performance, 25,* 482-503.

Ratcliff, R. (1988). A note on mimicking additive reaction time models. *Journal of Mathematical Psychology, 32,* 192-204.

Rayner, K. (1979). Eye guidance in reading: Fixation locations within words. *Perception, 8,* 21-30.

Rayner, K., McConkie, G. W., & Zola, D. (1980). Integrating information across eye movements. *Cognitive Psychology, 12,* 206-26.

Rayner, K., & Morris, R. K. (1992). Eye movement control in reading: Evidence against semantic preprocessing. *Journal of Experimental Psychology: Human Perception and Performance, 18,* 163-172.

Rayner, K., Sereno, S. C., & Raney, G. E. (1996). Eye movement control in reading: A comparison of two types of models. *Journal of Experimental Psychology: Human Perception and Performance, 22,* 1188-1200.

Rey, A., Chesnet, D., Bijeljac-Babic, R., Grainger, J., & Jacobs, A. M. (under revision). *A lightness increasing paradigm (LIP).* Manuscript submitted for publication.

Rey, A., Jacobs, A. M., Schmidt-Weigand, F., & Ziegler, J. C. (1998). A phoneme effect in visual word recognition. *Cognition, 68,* 41-50.

Rey, A., Ziegler, J. C., & Jacobs, A. M. (2000). Graphemes are perceptual reading units. *Cognition, 75,* B1-B12.

Richter, K. (1999). *A functional unit model of visual word recognition.* Unpublished Diploma-Thesis, Philipps-University Marburg, Marburg, Germany.

Roberts, S., & Sternberg, S. (1993). The meaning of additive reaction time effects: Test of three alternatives. In D. E. Meyer & S. Kornblum (Eds.), *Attention and performance XIV: Synergies in experimental psychology, artificial intelligende and cognitive neuroscience* (pp. 611-653). Cambridge, MA: MIT Press.

Rudnicky, A. I., & Kolers, P. A. (1984). Size and case of type as stimuli in reading. *Journal of Experimental Psychology: Human Perception and Performance, 10,* 231-249.

Rumelhart, D. E., & McClelland, J. L. (1982). An interactive activation model of context effects in letter perception: Part 2. The contextual enhancement effect and some tests and extensions of the model. *Psychological Review, 89,* 60-94.

Rumelhart, D. E., & Siple, P. (1974). Process of recognizing tachistoscopically presented words. *Psychological Review, 81,* 99-118.

Sanocki, T. (1987). Visual knowledge underlying letter perception: Font-specific, schematic tuning. *Journal of Experimental Psychology: Human Perception and Performance, 13,* 267-278.

Sanocki, T. (1988). Font regularity constraints on the process of letter recognition. *Journal of Experimental Psychology: Human Perception and Performance, 14,* 472-480.

Schmidt-Weigand, F. (1999). *Analyses of Single Words' Response Time Distributions: Applying Alternative Methods to Common Questions of Visual Word Recognition.* Unpublished Diploma-Thesis: Philipps-University Marburg, Marburg, Germany.

Schmidt-Weigand, F., Rey, A., Nuerk, H.-C., Graf, R., Jacobs, A. M., & Van Orden, G. C. (1998). Zum Verhältnis von Reaktionszeiten und Worteigenschaften in der visuellen Worterkennung [On the relation of reaction time and word attributes in visual word recognition]. In H. Lachnit, F. Rösler, & A. M. Jacobs (Eds.), *Experimentelle Psychologie: Abstracts der 40. Tagung experimentell arbeitender Psychologen* (p. 355). Lengerich: Pabst Science Publishers.

Sears, C. R., Hino, Y., & Lupker, S. J. (1995). Neighborhood size and neighborhood frequency effects in word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 21,* 876-900.

Seidenberg, M. S., & Plaut, D. C. (1998). Evaluating word-reading models at the item level: Matching the grain of theorey and data. *Psychological Science, 9,* 234-237.

Seidenberg, M. S., Waters, G. S., Barnes, M. A., & Tanenhaus, M. K. (1984). When does irregular spelling or pronunciation influence word recognition? *Journal of Verbal Learning and Verbal Behavior, 23,* 383-404.

Smith, F. (1969). Familiarity of configuration vs. discriminability of features in the visual identification of words. *Psychnomic Science, 14,* 261-263.

Smith, F., Lott, D., & Cronnel, B. (1969). The effect of type size and case alternation on word identification. *American Journal of Psychology, 82,* 248-253.

Snodgrass, J. G., & Mintzer, M. (1993). Neighborhood effects in visual word recognition: Facilitatory or inhibitory? *Memory and Cognition, 21,* 247-266.

Snodgrass, J. G., & Poster, M. (1992). Visual-word recognition thresholds for screen-fragmented names of the Snodgrass and Vanderwart pictures. *Behavior Research Methods, Instruments, and Computers, 24,* 1-15.

Spieler, D. H., & Balota, D. A. (1997). Bringing computational models of word naming down to the item level. *Psychological Science, 8,* 411-416.

Springer, S. P., & Deutsch, G. (1998). *Left brain, right brain: Perspectives from cognitive neuroscience* (5th ed.) New York, NY: Freeman.

Steiger J. H. (1993). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research, 25,* 173-180.

Sternberg, S. (1969). The discovery of processing stages: Extensions of Donders' method. *Acta Psychologia, 30,* 276-315.

Stone, G. O., Vanhoy, M., & Van Orden, G. C. (1997). Perception is a two-way street: Feedforward and feedback phonology in visual word recognition. *Journal of Memory and Language, 36,* 337-359.

Stone, G. O., & Van Orden, G. C. (1993). Strategic control of processing in word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 19,* 744-774.

Treiman, R., Mullennix, J., Bijeljac-Babic, R., & Richmond-Welty, E. D. (1995). The special role of rimes in the description, use, and acquisition of English orthography. *Journal of Experimental Psychology: General, 124,* 107-136.

Underwood, G., Clews, S., & Everatt, J. (1990). How do readers know where to look next? Local information distributions influence eye fixations. *Quarterly Journal of Experimental Psychology, 42A,* 39-65.

Vaina, L. M., Sundareswaran, V., & Harris, J. G. (1995). Learning to ignore: Psychophysics and computational modeling of fast learning of direction in noisy motion stimuli. *Cognitive Brain Research, 2,* 155-163.

Van Heuven, W. J. B., Dijkstra, T., & Grainger, J. (1998). Orthographic Neighborhood Effects in Bilingual Word Recognition. *Journal of Memory and Language, 39,* 458-483.

Van Orden, G. C., & Goldinger, S. D. (1994). Interdependence of form and function in cognitive systems explains perception of printed words. *Journal of Experimental Psychology: Human Perception and Performance, 20,* 1269-91.

Van Zandt, T., & Ratcliff, R. (1995). Statistical mimicking of reaction time data: Single-process models, parameter variability, and mixtures. *Psychonomic Bulletin and Review, 2,* 20-54.

Vitu, F., & O'Regan, J. K. (1988). Effect of parafoveal preprocessing and reading rythm on optimal landing position in words of different length and frequency. In G. Luer, U. Lass, & J. Schallo-Hoffmann (Eds.), *Eye movement research: Physiological and psychological aspects* (pp. 286-292). Toronto: Hogrefe.

Vitu, F., O'Regan, J. K., Inhoff, A. W., & Topolski, R. (1995). Mindless reading: Eye-movement characteristics are similar in scanning letter strings and reading texts. *Perception and Psychophysics, 57,* 352-64.

Vitu, F., O'Regan, J. K., & Mittau, M. (1990). Optimal landing position in reading isolated words and continuous text. *Perception and Psychophysics, 47,* 583-600.

Vonk, W., Radach, R., & van Rijn, H. (2000). Eye guidance and the saliency of word beginnings in reading text. In A. Kennedy, R. Radach, D. Heller, & J. Pynte (Eds.), *Reading as a perceptual process* (pp. 269-299). Oxford: Elsevier.

Weekes, B. S. (1997). Differential effects of number of letters on word and nonword naming latency. *Quarterly Journal of Experimental Psychology, 50A,* 439-456.

Weekes, N. Y., Capetillo-Cunliffe, L., Rayman, J., Iacoboni, M., & Zaidel, E. (1999). Individual differences in the hemispheric specialization of dual route variables. *Brain and Language, 67,* 110-133.

Wheeler, D. D. (1970). Processes in word recognition. *Cognitive Psychology, 1,* 59-85.

Wilding, J. M. (1988). The interaction of word frequency and stimulus quality in the lexical decision task. *Quarterly Journal of Experimental Psychology, 40A,* 757-770.

Wimmer, H., & Goswami, U. (1994). The influence of orthographic consistency on reading development: Word recognition in English children. *Cognition, 51,* 91-103.

Ziegler, J. C., Montant, M., & Jacobs, A. M. (1997). The feedback consistency effect in lexical decision and naming. *Journal of Memory and Language, 37,* 533-554.

Ziegler, J. C., & Perry, C. (1998). No more problems in Coltheart's neighborhood: Resolving neighborhood conflicts in the lexical decision task. *Cognition, 68,* 853-862.

Ziegler, J. C., Rey, A., & Jacobs, A. M. (1998). Simulating individual word identification thresholds and errors in the fragmentation task. *Memory and Cognition, 26,* 490-501.

Zorzi, M., Houghton, G., & Butterworth, B. (1998). Two routes or one in reading aloud: A connectionist dual-process model. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 1131-1161.

VII. Appendices

Appendix A:

Stimuli of Experiment 1: Nouns: Note that due to the Latin Square Design and within-participant presentation of different word forms, i.e., repetitions, individual item statistics are not meaningful and are hence not reported in Experiment 1 and 2 (see text for details). The statistical analyses for Experiment 1 and all other experiments were performed with STATISTICA or SPSS 9.0.

Stimuli	logF*	Stimuli	log*
Abend	2.12	Mitte	2.13
Angst	2.01	Motor	1.98
Beruf	1.94	Musik	2.08
Blatt	1.87	Nacht	2.21
Blick	2.25	Natur	2.08
Boden	2.18	Opfer	2.05
Brief	2.15	Osten	1.95
Ernst	2.29	Pferd	1.82
Essen	2.21	Prinz	1.81
Fahrt	1.81	Radio	1.79
Farbe	1.84	Regel	1.93
Firma	2.10	Reich	2.16
Folge	2.22	Reihe	2.19
Geist	2.06	Rolle	2.23
Hotel	2.07	Sorge	2.04
Insel	1.85	Sport	1.90
Junge	2.12	Stand	1.90
Kunst	2.05	Stern	1.85
Lager	1.86	Thema	2.13
Licht	2.04	Tisch	2.00
Liebe	2.06	Titel	1.93
Linie	2.04	Union	1.86
Markt	2.02	Waffe	1.91
Masse	1.84	Wagen	2.32
Menge	1.94	Wesen	2.04

* logF always refers to the logarithm to the base 10 of (total word frequency / million) +1 (1 is added to prevent negative logarithmic values for frequencies smaller than 1).

Appendix B:

Stimuli of Experiment 2: Non-Nouns: Note that due to the Latin Square Design and within participant presentation of different word forms, i.e., repetitions individual item statistics are not meaningful and are hence not reported in Experiment 1 and 2 (see text for details).

Stimuli	logF	Stimuli	logF
allzu	1.67	hinzu	1.78
bevor	2.01	holen	2.08
blind	1.47	hören	2.58
blond	1.36	innen	1.41
brach	1.35	jener	2.57
daher	2.27	knapp	1.92
dahin	1.79	leise	1.53
darin	2.26	links	1.87
darum	2.23	malen	1.34
davon	2.54	manch	2.25
davor	1.47	meist	1.60
desto	1.33	nackt	1.43
drauf	1.45	nahen	1.36
eilen	1.21	offen	2.27
eilig	1.23	rasch	1.94
enorm	1.25	sanft	1.45
exakt	1.47	somit	1.53
genau	2.51	sonst	2.28
genug	2.20	sowie	2.57
gerne	1.75	steil	1.23
glatt	1.50	stets	2.02
heben	2.06	stumm	1.27
herab	1.29	teuer	1.84
heran	1.54	total	1.77
herum	1.75	umher	1.20

Appendix C1:

High- and low-frequent nouns presented in Experiment 3. Nouns that were faster recognized in IUC form are marked with an ”*”.

Stimuli	logF	RT_IUC	RT_ALC	Stimuli	logF	RT_IUC	RT_ALC
Abend	2.12	558	515	Altar	0.95	623	612
Alter	1.73	547	524	Anker*	0.62	534	569
Armee	1.84	593	584	Anmut*	0.65	620	740
Autor*	1.79	514	615	Biene	0.90	577	531
Basis*	1.75	592	627	Birke*	0.56	571	602
Beruf*	1.94	535	560	Erker	0.37	712	683
Blick*	2.25	537	634	Esche*	0.37	685	712
Boden*	2.18	568	571	Fakir	0.07	724	704
Dauer	1.74	542	525	Faser	0.90	672	663
Feuer	1.74	545	524	Harfe*	0.26	641	676
Firma	2.10	623	577	Hauch*	0.91	551	625
Front*	1.93	592	609	Henne*	0.85	614	657
Hotel*	2.07	549	587	Joker*	0.07	571	616
Junge*	2.12	504	542	Kerbe*	0.30	658	735
Krieg*	2.50	550	602	Keule	0.48	548	546
Krise*	1.80	568	633	Komik*	0.62	563	612
Leute*	2.26	503	574	Koran*	0.22	671	711
Markt	2.02	613	531	Lehne*	0.54	636	716
Menge*	1.94	556	623	Limit	0.34	710	706
Meter*	2.32	565	576	Molch*	0.50	611	704
Monat*	2.52	507	555	Motel*	0.37	639	770
Motor*	1.98	551	575	Narbe*	0.98	507	556
Musik*	2.08	556	560	Neige*	0.50	602	691
Nacht*	2.21	520	578	Nomen*	0.07	727	903
Natur*	2.08	526	586	Notar*	0.99	528	666
Opfer*	2.05	509	601	Nutte*	0.07	601	670
Organ	1.75	594	518	Optik	0.86	577	512
Platz	2.41	585	546	Pfand	0.34	548	534
Preis*	2.42	573	596	Poker	0.22	664	624
Prinz*	1.81	499	590	Quark*	0.22	561	714
Roman*	1.77	527	576	Satan*	0.96	568	724
Sache*	2.42	574	608	Seife*	0.88	539	546
Seite	2.80	558	535	Skalp*	0.22	629	806
Stadt*	2.61	530	561	Spule*	0.87	569	637
Strom*	1.74	540	552	Spurt	0.64	644	618
Stufe*	1.72	571	616	Stier*	0.98	515	573
Titel*	1.93	526	639	Tabak*	0.97	600	639
Vater*	2.37	526	574	Walze*	0.62	590	620
Waffe	1.91	566	551	Wespe	0.50	631	629
Zweck*	1.75	524	541	Wulst	0.22	687	679

Appendix C2:

High- and low-frequent non-nouns presented in Experiment 3. Non-nouns that were faster recognized in ALC form are marked with an ”*”. Non-nouns that were eliminated because of an accuracy of lower than .67 are marked with an ”#”.

Stimuli	logF	RT_IUC	RT_ALC	Stimuli	logF	RT_IUC	RT_ALC
aktiv	1.97	565	537	adlig	0.58	727	647
bevor*	2.01	563	586	affig	0.12	735	658
breit*	2.00	553	627	anbei*	0.07	579	627
daher*	2.27	572	586	bange	0.65	625	601
dahin*	1.79	560	590	blank	1.00	570	561
daran	2.36	642	572	dumpf*	1.02	605	636
darin*	2.26	508	608	duzen*	0.62	704	745
darum*	2.23	525	559	eitel*	0.78	568	581
dicht*	1.76	586	600	empor*	1.04	612	652
einig*	2.39	565	572	emsig	0.60	613	603
falls*	1.71	506	552	erdig	0.07	618	613
fremd	1.73	579	554	famos*	0.37	650	701
genau	2.51	554	531	fatal*	0.78	594	681
genug*	2.20	550	570	feige	0.58	589	578
gerne*	1.75	582	654	fidel*	0.81	665	749
herum*	1.75	602	606	final	0.18	706	681
hinzu*	1.78	535	581	flink*	0.78	546	639
holen*	2.08	606	616	flugs#	0.37	#	#
indem*	2.06	610	659	frech*	0.97	567	633
klein*	2.76	597	604	grell*	0.83	580	684
knapp	1.92	581	581	infam#	0.37	#	#
lange*	1.95	531	605	irrig	0.58	764	636
lesen*	2.20	526	582	kauen	0.88	578	569
links	1.87	569	523	krumm*	0.95	570	603
neben	2.41	585	574	laben*	0.12	707	733
offen*	2.27	573	588	madig	0.12	749	680
rasch*	1.94	562	619	mimen#	0.43	#	#
ruhig	2.15	636	560	mobil*	0.54	540	599
sogar	2.46	547	532	nagen	0.30	745	641
sonst*	2.28	536	562	nasal	0.07	809	780
sowie*	2.57	549	588	platt	0.73	631	570
statt*	2.28	559	570	plump*	0.77	654	696
stets*	2.02	580	748	reell#	0.54	#	#
teuer*	1.84	568	662	rings*	0.81	667	689
total	1.77	554	544	senil	0.12	708	696
warum*	2.34	520	591	tosen*	0.30	709	729
weder*	2.09	578	625	trist*	0.40	602	653
wobei	1.92	569	569	tuten*	0.45	729	736
zumal	1.72	643	619	ungut	0.48	598	597
zuvor*	1.90	548	583	vorab*	0.65	576	616

Appendix C3:

Nonwords derived from high- and low-frequent nouns presented in Experiment 3. Nonwords slower rejected in IUC form are marked with an ”*”.

Nonwords derived from high-frequent nouns				Nonwords derived from low-frequent nouns			
Stimuli	logF	RT_IUC	RT_ALC	Stimuli	logF	RT_IUC	RT_ALC
Aberd	0.00	657	680	Ankor	0.00	625	685
Alted	0.00	664	675	Anmet	0.00	674	690
Armie*	0.00	731	701	Antar	0.00	650	653
Autol*	0.00	690	630	Bienu	0.00	637	664
Benuf	0.00	667	713	Birka*	0.00	738	690
Bliek*	0.00	693	630	Erher	0.00	722	740
Bofen	0.00	606	638	Escke	0.00	663	762
Bosis*	0.00	685	655	Fasar*	0.00	670	619
Dauar*	0.00	713	712	Fokir*	0.00	653	593
Feuar	0.00	661	772	Haich	0.00	572	627
Firsa*	0.00	620	605	Harfa*	0.00	751	666
Frort	0.00	659	665	Henno*	0.00	647	557
Hogel*	0.00	675	575	Joler*	0.00	667	636
Jurge*	0.00	687	638	Karan	0.00	622	649
Krigg	0.00	565	618	Kerse	0.00	681	699
Kripe*	0.00	791	644	Keuge*	0.00	710	695
Leete	0.00	559	664	Korik*	0.00	634	632
Malkt*	0.00	657	587	Lehme	0.00	684	788
Mator	0.00	645	654	Libit	0.00	620	639
Metes*	0.00	655	624	Molck*	0.00	683	589
Mogat*	0.00	669	655	Monel*	0.00	683	582
Monge	0.00	627	630	Narbu*	0.00	654	592
Musim*	0.00	709	659	Noige	0.00	591	678
Netur*	0.00	622	593	Nosar*	0.00	682	580
Nocht	0.00	619	627	Nuhte	0.00	619	648
Olfer*	0.00	627	584	Nutto	0.00	677	696
Ordan*	0.00	718	700	Oftik	0.00	662	701
Pfatz*	0.00	733	704	Porer	0.00	664	673
Prein*	0.00	690	664	Psand	0.00	602	625
Prenz*	0.00	625	588	Quank*	0.00	678	675
Reman	0.00	641	671	Saman*	0.00	678	628
Sacha*	0.00	625	579	Spalp	0.00	656	691
Seita	0.00	651	686	Spula	0.00	639	640
Spadt*	0.00	774	691	Spurg	0.00	672	702
Spufe*	0.00	675	659	Stife*	0.00	774	703
Strog	0.00	662	701	Stirr*	0.00	786	779
Tisel	0.00	607	657	Tebak*	0.00	732	637
Vatur*	0.00	620	606	Welze*	0.00	674	662
Waffa*	0.00	663	635	Wespo	0.00	619	664
Zwech*	0.00	683	660	Wulft*	0.00	648	576

Appendix C4:

Nonwords derived from high- and low-frequent non-nouns presented in Experiment 3. Nonwords slower rejected in ALC form are marked with an ”*”.

Nonwords derived from high-frequent nouns				Nonwords derived from low-frequent nouns			
Stimuli	logF	RT_IUC	RT_ALC	Stimuli	logF	RT_IUC	RT_ALC
amtiv	0.00	637	675	adlin*	0.00	757	606
bovor	0.00	684	724	afeig	0.00	629	694
brein	0.00	663	725	anbri	0.00	580	613
dahan*	0.00	688	675	banee	0.00	569	620
daheb	0.00	649	652	blang*	0.00	771	645
daram*	0.00	781	724	duben	0.00	651	694
dasin*	0.00	759	724	dummf	0.00	779	784
dehin*	0.00	720	614	emfor*	0.00	644	639
dieht	0.00	632	637	emsag	0.00	614	632
eineg*	0.00	665	664	erdil	0.00	629	659
folls	0.00	621	634	estel	0.00	592	648
fregd	0.00	579	617	famis	0.00	600	646
ganug	0.00	703	710	fatar	0.00	696	769
gedau	0.00	674	674	fiden*	0.00	681	670
gerna	0.00	579	624	fival*	0.00	636	596
heram*	0.00	654	652	flech*	0.00	685	637
hilzu*	0.00	638	638	flirk	0.00	718	730
holes	0.00	672	687	flugt	0.00	685	714
indom	0.00	599	662	frige	0.00	637	758
kleig*	0.00	694	693	grull*	0.00	663	645
knapf*	0.00	777	750	ireig*	0.00	686	669
langa	0.00	623	645	irfam	0.00	616	670
lefen*	0.00	640	599	keuen	0.00	690	720
linps	0.00	597	702	kremm*	0.00	635	633
neden	0.00	587	689	laban*	0.00	648	644
offon*	0.00	679	673	mabig	0.00	631	661
resch	0.00	652	712	memen	0.00	653	726
rufig*	0.00	749	742	monil	0.00	629	642
sanst*	0.00	751	688	nagun*	0.00	667	560
sawie*	0.00	671	609	nesal	0.00	602	619
soger*	0.00	701	688	plemp*	0.00	685	608
spets	0.00	634	635	pratt*	0.00	756	700
stutt*	0.00	668	642	raell*	0.00	659	630
tetal	0.00	631	659	rinns	0.00	591	756
teuel	0.00	654	692	sonil*	0.00	674	532
warud	0.00	605	644	toset*	0.00	605	605
wedar	0.00	646	648	trisp*	0.00	627	600
wobbi*	0.00	600	578	tutan*	0.00	687	672
zubal	0.00	655	723	unget	0.00	626	634
zuvog*	0.00	664	663	vorag*	0.00	744	728

Appendix D:

Stimuli of Experiment 4: Because due to the Latin square design items were not presented under identical conditions to different participants, the prerequisites for item analysis are not fulfilled. Item accuracy values are subject to a participant by viewing position interaction and, hence, not reliable. Therefore, these accuracy values are not given.

3 Letters		5 Letters		7 Letters	
low Freq	**high Freq**	**low Freq**	**high Freq**	**low Freq**	**high Freq**
AHN	ARM	BESEN	BEIDE	BALANCE	ANDERER
BUG	BEI	BIEST	DABEI	BERSTEN	AUFGABE
DIA	BIS	DOLCH	DURCH	BRESCHE	BLEIBEN
DUO	EHE	ERKER	EIGEN	EINFALT	BRINGEN
DUR	ENG	FERNE	EINEN	HEILBAR	DEUTSCH
FAD	GUT	FIBEL	ERSTE	HINTERN	ERGEBEN
FEE	HER	GICHT	ETWAS	KITZELN	ERHEBEN
FIX	HIN	HAUBE	FRAGE	KRALLEN	GERICHT
GEL	ICH	HUMUS	GEHEN	KULLERN	GESICHT
HAB	IHR	LIMIT	GRUND	PINSELN	GESTERN
HAI	NEU	LOCKE	HEUEN	PRAHLEN	GLAUBEN
ION	NIE	LURCH	HEUTE	PRELLEN	HANDELN
JOD	NUN	NIETE	IMMER	REIBUNG	MILLION
JUX	OFT	ORKAN	KLEIN	REIZUNG	PROZENT
KLO	ORT	PARAT	KRAFT	RIESELN	RICHTEN
KOT	POL	PFOTE	RECHT	SCHABEN	RICHTIG
LUG	TOD	PUDER	SCHON	SCHLAPP	SCHRITT
ODE	TUN	QUERE	SEITE	SCHLIFF	SCHWARZ
SET	UHR	RITZE	SOLCH	STIEBEN	SONDERN
UHU	UNS	RODEN	STAAT	STOLLEN	SPIELEN
ULK	VOR	SCHAL	STADT	TRUNKEN	STEIGEN
VIA	WAS	SEKTE	STARK	VORHAND	STELLEN
ZEH	WEG	SIPPE	UNSER	WEGFALL	STIMMEN
ZIG	WIE	TOLLE	UNTER	WETTERN	WICHTIG
ZWO	ZUG	WISCH	WENIG	ZWINGER	ZEITUNG

Appendix E:

Stimulus Level of Correct Responses and Accuracy in Experiment 5.

	High Frequency			Low Frequency		
	Stimuli	**Level**	**Acc**	**Stimuli**	**Level**	**Acc**
Low LCI	ARZT	4.75	0.80	GANS	3.50	1.00
	BANK	4.38	0.80	KAHN	4.44	0.90
	BAUM	3.44	0.90	KAUZ	4.67	0.90
	DANK	4.30	1.00	KOKS	4.63	0.80
	GAST	3.80	1.00	KRAN	5.30	1.00
	HAAR	5.00	0.90	MAAT	5.63	0.80
	KAUM	4.89	0.90	MARS	3.11	0.90
	KURS	5.33	0.90	MOST	4.70	1.00
	MANN	3.60	1.00	RAHM	4.63	0.80
	MARK	4.00	1.00	RANK	5.22	0.90
	RANG	4.44	0.90	RAST	4.63	0.80
	SATZ	4.50	0.80	SAAT	3.56	0.90
	SINN	4.40	1.00	SAUM	4.40	1.00
	SOHN	5.22	0.90	SIMS	4.80	1.00
	WAHR	5.20	1.00	SKAT	4.67	0.90
	WANN	4.89	0.90	TRAM	4.57	0.70
	WARM	3.50	1.00	WAHN	4.33	0.90
	WERK	5.11	0.90	WAMS	5.30	1.00
	ZAHN	5.63	0.80	WURM	4.56	0.90
	ZWAR	5.50	1.00	ZAUM	4.57	0.70
High LCI	BALL	5.33	0.90	BEET	6.88	0.80
	BIER	5.43	0.70	BEIL	5.70	1.00
	DICH	5.78	0.90	BREI	6.00	0.60
	DORF	4.40	1.00	DUFT	5.67	0.90
	DREI	5.00	1.00	FEIG	5.89	0.90
	ECHT	5.00	0.90	FEIL	5.60	0.50
	FALL	5.40	1.00	FELL	5.43	0.70
	FELD	4.88	0.80	FILZ	5.80	1.00
	FREI	6.10	1.00	FLOH	5.38	0.80
	GOLD	4.80	1.00	FLOR	6.20	0.50
	HELD	5.13	0.80	HIEB	6.40	0.50
	HELL	4.71	0.70	HOLD	5.90	1.00
	KIEL	5.56	0.90	KEIL	6.33	0.90
	LAUF	5.83	0.60	LAIB	5.78	0.90
	LEER	6.00	0.80	LECK	6.40	1.00
	LIED	5.50	0.80	LIFT	6.20	1.00
	TEIL	5.67	0.90	NEID	5.70	1.00
	VIEL	5.63	0.80	PILZ	5.40	1.00
	VOLL	5.43	0.70	PULT	5.22	0.90
	ZIEL	4.90	1.00	VOLT	4.90	1.00

Appendix F:

Letter confusability indices (LCI) of the single letters in Experiment 4.

Letters	Raw Conf. Index	z-Normalization	LCI = Φ(-z)
A	575	0.66	0.25
B	435	-0.51	0.70
C	330	-1.38	0.92
D	398	-0.81	0.79
E	371	-1.04	0.85
F	337	-1.32	0.91
G	485	-0.09	0.54
H	513	0.14	0.44
I	342	-1.28	0.90
J	493	-0.02	0.51
K	598	0.85	0.20
L	297	-1.65	0.95
M	696	1.66	0.05
N	568	0.60	0.27
O	459	-0.31	0.62
P	374	-1.01	0.84
Q	478	-0.15	0.56
R	525	0.24	0.41
S	696	1.66	0.05
T	486	-0.08	0.53
U	468	-0.23	0.59
V	399	-0.80	0.79
W	740	2.02	0.02
X	651	1.29	0.10
Y	563	0.56	0.29
Z	619	1.02	0.15

The raw letter confusability index was computed as follows. Each letter was presented 100 times in a pilot study. For each presentation, each letter received the score Conf = (9- level of correct recognition) if it was responded correctly and Conf = 0 if it was responded incorrectly. These scores were summed up over 100 presentations per participant.

The LCI for the single letters used in this study was the normal distribution function over the negative standard-normalized z-value of this raw letter confusability index. The raw value was z-normalized over letters rather than words to keep the LCI of any given word independent of word selection in a study. The negative z-value was taken to index LCI values congruent to the term confusability (high LCI values represent high confusability). The Φ-values were taken, to prevent that one single very low confusable letter (e.g., W) determines a low confusability value of an otherwise average confusable word. The LCI for each word was computed as the average of the LCIs for its constituent letters.

Appendix G: Stimuli of Experiment 6:

Negative N-Grad					
low Freq	Accuracy	Middle Freq	Accuracy	High Freq	Accuracy
BEBEN	0.40	FEDER	0.53	DAUER	0.87
DEGEN	0.73	JAGEN	0.87	DEINE	0.67
DELLE	0.33	KABEL	0.80	DICHT	0.73
FEGEN	0.20	KADER	0.53	GEGEN	0.67
GASSE	0.87	KISTE	0.80	GEIST	0.87
GATTE	0.47	KOHLE	0.87	HEBEN	0.53
GICHT	0.33	LACHE	0.47	JEDER	0.80
HAGEN	0.67	LEDER	0.80	LEBEN	0.87
HEGEN	0.47	LINDE	0.87	LEGEN	0.60
HELLE	0.20	LIPPE	0.80	LICHT	1.00
HIPPE	0.40	LOSEN	0.47	MACHT	0.73
KIPPE	0.47	MAGEN	0.93	MASSE	0.80
LATTE	0.73	NONNE	0.87	MAUER	0.80
MATTE	0.80	PILLE	0.87	NACHT	1.00
NAGEN	0.47	RACHE	0.60	NEBEN	0.60
PACHT	0.33	RASSE	0.73	NICHT	0.87
PELLE	0.33	SAUER	0.73	REGEN	0.93
RAGEN	0.27	SEGEN	0.93	ROLLE	0.80
RIPPE	0.53	SICHT	0.80	SACHE	0.73
SACHT	0.40	TAGEN	0.60	SAGEN	0.53
SIPPE	0.47	TASSE	0.80	SONNE	0.93
WATTE	0.73	WELLE	0.53	TONNE	0.93
WICHT	0.87	WOLLE	0.93	WAGEN	0.87
WIPPE	0.73	ZEILE	0.27	WEGEN	0.53
YACHT	0.93	ZELLE	0.67	WILLE	0.27
Positive N-Grad					
low Freq	Accuracy	Middle Freq	Accuracy	High Freq	Accuracy
ALTAR	0.33	AGENT	0.73	AKTIE	0.13
BLUSE	0.47	ALLEE	0.20	AKTIV	0.87
BRAUE	0.07	BETEN	0.27	BRIEF	0.80
FOLIE	0.40	BLOCK	0.93	EINER	0.20
HINDU	0.40	BRAUN	0.93	FALLS	0.67
INDES	0.53	BRAUT	0.80	GENUG	0.73
INDEX	0.67	EIFER	0.13	INDEM	0.40
KRAUL	0.40	GEBET	0.47	KLEIN	0.73
KRUME	0.13	HAFEN	0.53	KRAFT	0.53
MODUL	0.53	KRANZ	0.93	LINKE	0.40
QUARK	0.87	KURVE	0.93	LINKS	0.93
QUART	0.20	NAGEL	1.00	RADIO	0.93
QUARZ	0.93	SCHAF	0.60	SCHON	0.93
QUAST	0.60	SCHAR	0.20	SPIEL	0.93
REIBE	0.53	SCHAU	0.73	SPORT	0.87
SCHAH	0.07	SCHUH	0.93	STAAT	0.47
SCHAL	0.67	STARR	0.20	STADT	0.80
SCHAM	0.67	STEIF	0.47	STARK	0.87
SCHUB	0.60	STEIL	0.20	START	0.60
SENSE	0.60	STUBE	0.67	STATT	0.40
SPORE	0.07	STURM	0.87	STEIN	0.60
SPORN	0.47	STURZ	0.73	STERN	0.73
STEIG	0.40	STUTE	0.40	STUFE	0.47
STOLA	0.53	WEITE	0.13	UNTEN	0.73
TROTT	0.33	ZWEIG	0.87	UNTER	0.67

Note, that due to the Latin Square Design the stimuli are not presented under identical conditions to the different participants. Therefore, an item ANOVA is usually not performed (cf. O'Regan & Jacobs, 1992).

Appendix H:

Letter confusion matrix used to determine N-conf for each word. The letters in the rows depict the presented letters, the letters in the columns depict the letter responses. The letter confusion matrix depicts, how often each letter was confused with each other letters. Note, that exchanges are asymmetrical. So, a presented “A” was once wrongly answered as “B”, but a presented B was never wrongly answered as “A”.

Note that for some letters less than 100 responses were given due to null responses when nothing was typed in. In these cases the relative (percent) confusion value was used. N-conf for each word was simply computed as the cumulative frequencies of those (percent) values for the very letter exchanges critical for neighbor creation.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	
A	**70**	1	0	0	0	1	1	1	0	0	8	0	4	4	0	0	0	2	0	0	0	0	3	4	0	0	99
B	0	**41**	2	10	5	4	0	6	1	0	2	1	0	0	3	12	0	10	0	2	1	0	0	0	0	0	100
C	0	0	**60**	2	0	0	1	0	0	0	0	0	0	0	27	0	1	0	8	0	1	0	0	0	0	0	100
D	0	1	3	**58**	0	1	0	4	1	0	1	4	3	7	5	4	0	2	0	2	3	0	1	0	0	0	100
E	0	2	1	1	**37**	34	1	1	1	0	2	9	0	0	0	4	0	4	0	1	1	0	0	0	0	0	99
F	0	1	0	2	16	**55**	3	1	10	0	1	2	0	0	0	0	0	2	0	1	5	0	0	0	1	0	100
G	0	0	22	2	1	0	**39**	0	0	0	0	0	0	1	9	0	3	1	21	0	0	1	0	0	0	0	100
H	4	0	0	1	7	4	1	**38**	3	1	1	2	3	5	0	3	0	5	2	1	18	0	0	0	0	0	99
I	0	1	0	5	7	2	1	1	**59**	0	2	7	1	1	3	1	0	0	0	3	2	1	3	0	0	0	100
J	0	0	1	0	0	0	0	0	23	**54**	0	2	0	0	0	0	3	0	4	2	8	1	0	0	2	0	100
K	0	0	0	0	0	2	1	0	1	0	**79**	2	1	1	0	0	0	10	0	1	0	0	0	1	0	0	99
L	0	0	0	2	16	2	0	0	27	5	1	**29**	1	0	0	5	0	1	2	2	1	0	0	1	2	2	99
M	0	0	0	0	0	0	0	1	0	0	0	0	**95**	3	0	0	0	0	0	0	0	0	1	0	0	0	100
N	0	0	0	1	0	1	2	0	0	2	2	0	14	**72**	0	0	0	1	1	0	2	1	1	0	0	0	100
O	0	0	14	0	1	0	1	0	0	1	0	0	0	0	**73**	0	2	0	6	0	1	1	0	0	0	0	100
P	0	5	0	2	4	7	0	6	4	0	1	1	0	0	1	**58**	0	10	0	1	0	0	0	0	0	0	100
Q	0	0	3	0	0	0	0	0	0	0	1	0	0	0	27	0	**65**	3	0	0	0	0	0	0	0	0	99
R	0	2	0	6	5	1	0	1	3	0	9	0	2	1	0	6	2	**56**	0	5	0	0	0	0	1	0	100
S	0	2	1	1	0	0	2	0	0	0	0	0	0	1	1	0	1	1	**90**	0	0	0	0	0	0	0	100
T	0	0	3	0	6	5	1	0	28	3	0	3	1	0	1	0	0	2	1	**42**	3	0	0	0	0	1	100
U	0	3	9	3	2	0	1	1	4	11	0	1	0	1	8	0	0	0	1	0	**55**	0	0	0	0	0	100
V	1	0	0	0	0	1	0	0	0	0	9	0	10	4	0	0	0	2	0	1	0	**42**	2	14	11	0	97
W	0	0	0	0	0	0	0	0	0	0	0	0	2	1	0	0	0	1	0	0	0	0	**96**	0	0	0	100
X	1	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	1	1	1	0	0	0	**87**	5	0	100
Y	0	1	0	0	0	0	0	0	0	0	3	0	0	0	0	0	1	0	1	6	0	1	0	18	**69**	0	100
Z	0	1	0	0	2	1	0	1	0	1	2	4	0	1	0	0	0	0	0	0	0	0	0	6	0	**81**	100
	76	61	119	96	109	121	55	62	165	78	128	67	137	103	158	93	78	114	138	71	101	48	107	131	91	84	2591

Appendix I: Stimuli of Experiment 7

Hermits	% Errors	Level	Low N-conf	% Errors	Level
ADLER	0.00	5.18	BARDE	18.18	5.00
CELLO	9.09	6.30	BESEN	18.18	4.44
DATIV	9.09	6.30	EIMER	9.09	5.30
DOGGE	0.00	5.55	EISIG	18.18	5.78
DROGE	0.00	4.91	FABEL	9.09	6.30
DUELL	9.09	6.20	FALTE	18.18	6.33
EDIKT	9.09	5.70	FEILE	0.00	6.55
FEHDE	18.18	6.00	GATTE	9.09	5.80
FLORA	18.18	5.00	GEIER	9.09	5.80
GELEE	18.18	6.78	GEIGE	9.09	5.30
IDYLL	18.18	6.33	JACKE	0.00	4.91
KALIF	27.27	6.88	JOLLE	27.27	6.63
KEFIR	0.00	6.27	KELLE	9.09	6.30
LARVE	0.00	5.55	KEULE	0.00	5.18
LOTSE	0.00	4.82	LAUBE	45.45	4.17
MOPED	0.00	5.18	LIEGE	0.00	6.36
NEFFE	18.18	5.33	LINSE	0.00	4.36
NELKE	9.09	4.90	LODEN	9.09	6.70
OPTIK	0.00	4.73	MEISE	9.09	5.00
ORGEL	0.00	5.00	MODEL	18.18	5.33
PEDAL	0.00	5.18	MOTTE	9.09	5.20
PFOTE	9.09	5.20	NADEL	9.09	4.40
PHOTO	0.00	5.00	NOBEL	18.18	5.67
POLKA	0.00	5.55	NUDEL	18.18	4.56
QUOTE	9.09	5.20	PUPPE	0.00	5.27
REEDE	27.27	7.50	REIFE	18.18	6.67
ROBBE	36.36	6.00	SEIFE	18.18	5.11
TIGER	0.00	5.00	SIPPE	0.00	5.45
TULPE	0.00	5.27	WIPPE	0.00	5.09
VIPER	0.00	5.27	ZECKE	9.09	4.40
Controls	**% Errors**	**Level**	**High N-conf**	**% Errors**	**Level**
ABGAS	9.09	5.40	BACKE	9.09	6.10
AHORN	0.00	3.82	BIEGE	18.18	7.00
ALARM	0.00	4.27	BINDE	0.00	6.27
ASPIK	9.09	5.00	BIRNE	0.00	5.27
ATLAS	9.09	4.00	BRISE	9.09	5.50
BIWAK	9.09	5.70	BULLE	27.27	6.13
BONUS	9.09	4.90	EICHE	27.27	6.00
DOGMA	0.00	4.73	EUTER	18.18	6.11
EKZEM	9.09	5.20	FELGE	45.45	6.67
GUMMI	0.00	4.82	FIBEL	27.27	5.88
HAREM	18.18	6.11	GARBE	18.18	6.22
HUMUS	0.00	4.27	GEHER	27.27	7.13
HUSAR	9.09	5.20	HAGEL	18.18	6.22
ISLAM	0.00	5.55	HOCKE	18.18	5.78
KAJAK	9.09	4.80	KEHLE	18.18	5.11
KASKO	0.00	6.00	KIPPE	18.18	3.89
KIOSK	0.00	4.09	KUPPE	27.27	5.63
KOMMA	18.18	4.44	LUNGE	0.00	4.91
LASSO	0.00	4.09	NIETE	9.09	6.00
LUXUS	27.27	4.38	PEGEL	9.09	5.20
MAGMA	0.00	4.73	PINTE	27.27	5.88
MIMIK	18.18	4.89	PISTE	9.09	5.30
MOKKA	9.09	5.00	PUDER	36.36	5.43
NOTAR	0.00	4.36	RAPPE	54.55	6.80
SAMBA	9.09	4.00	REIBE	18.18	5.78
SESAM	9.09	3.30	RILLE	36.36	6.43
TABAK	18.18	4.33	RUDEL	36.36	5.57
TORSO	0.00	5.64	SEGEL	9.09	4.40
UMZUG	0.00	4.27	WETTE	9.09	5.60
ZYSTE	27.27	4.13	WIEGE	9.09	5.40

Appendix J1: MatLab-Programs for the NAMWR: Master Program

```
---------------------------------
% program vpedissmaster
% hc 03 /2000

% Last revision hc 03 /2001
%
% the program needs a matrix called "ref" with the following format
%
%Column 1-5: empirical data in p-values
% Column 6: word length
% Column 7: Word Form Code: 1- All lower case; 2 - Initial Upper case; 3 - All Upper case
% Column 8: Syntactic class: 1 - Nouns; 0 - Non-nouns, value can be average in mixed lists
% Column 9: f-manipulation: 1-low; 2-middle; 3-high
% Column 10: Ngrad-Manipulation: 1: positive N-grad; -1: negative N-grad; 0: no N-grad manipulation
% Column 11: Log10(F/Mio+1)
% Column 12: meanNgrad: absolute value of mean ngrad
% Column 13: Syllables: Mean Number of Syllables
% Column 14: Mask: Number of Backward Mask ####s
% Column 15: N: Mean Number of Neighbors
% Columns 16-22: Mean Number of positional neighbors for positions 1-7

% The NAMWR of the thesis of HC Nuerk 2001 only uses the colums 1-6, 11, and 15-22.
% To fit more fine-grain data other word or presentation variables (e.g. Syllables, mask)
% may need to be considered. This is already prepared in the current model.

% The codes in column 9 and 10 are used by the program vpegraph which produces figures
% and titles

% definition of matrices
posmatrix = ref(:,1:5);
wl = ref(:,6);
wfcode = ref(:,7);
synclass = ref(:,8);
freqman= ref(:,9);
ngradman = ref(:,10);
log10freq = ref(:,11);
meanngrad = ref(:,12);
syllables = ref(:,13);
masknr = ref(:,14);
neighbors = ref(:,15);
posneighbors = ref(:,16:22);

% definition of manipulations

% the following manipulations and parameters need to be defined

% nfp: number of fixation positions
nfp=5;

% This is the output matrix
datasimmatrix=zeros(size(posmatrix,1),10);

% Because n curves shall now simultaneously be fitted,
% the number of fittings must be defined.
nrfittings=size(posmatrix,1);


% On this basis the whole matrix of viewing position curves is fitted.
wl =ref(:,6);
freq=ref(:,11);
```

```
nfp=nfp*ones(nrfittings,1);
ngrad=ref(:, 16:15+max(wl));

% x is the fixed parameter set. In the thesis, a variable y is taken instead of nfp.
% Because y is a function of nfp, this is mathematical the same.
x=[nfp,wl,freq ngrad];
fixedparas=x

% y is as x an index variable that is subject to change in the fitting process.
% It is NOT the same variable as in the thesis. Sorry.
y=posmatrix;
datasimmatrix(:,1:10)=vpefit2000matf(fixedparas,y);

% The output is saved in the file vpesims
save vpesims datasimmatrix -ascii;

% This is the output graph. It may need to be dropped or changed
% if the screen (pixel) graph differs strongly from mine

vpegraph
```

Appendix J2:

MatLab-Programs for the NAMWR: Program starting and evaluating simplex fitting

```
-----------------------------------------------------------
function [F] = vpefit2000matrixf(fixedparas,y)
% vpefitdata
% fits a number of vpe curves with and without parameter to empirical data
% C hcn 03/2000
% revised hcn 03/2001

% The function arguments are:
% fixedparas(1)=nfp: number of fixation positions (default is 5)
% fixedparas(2) wl: word length
% fixedparas(3)= freq: frequency
% fixedparas(4:10) = ngrad: neighborhood gradient

% The three free parameters are:
% p(1)=d: drop-off rate: The rate with which recogntion probability drops off:
%    influences height and steepness of the curve
% p(2)=a: asymmetry ratio: The higher "a" the more asyemmtric is the curve.
%   influences the asymmetry of the optimal viewing position
% p(3)=l lexical influence: at any given fixation probability l(1-p(f))*logF
% p(4)=r z-value of recognition likelihood on fixation for wl 5.


if size(y,2)==fixedparas(1,1)
   disp('now fitting')
   ymatemp= y;

      nrfittings=size(y,1);

      %Turn matrices into vector
   tempparas=fixedparas';
   fixedparavector= tempparas(:)';

   tempy=y';
   yvector= tempy(:)';



      % ******************* VPE-2000-SIMPLEX ********************

            start = [0.30 1.2 0.3 2];  % initial guessing vector

            % vpe2000 serves for data constructing
            Z = vpe2000mat(fixedparavector,start,nrfittings);
            p = fmins('vpe2000matf2',start,[0 0.01 0.01],[],fixedparavector,yvector);
            % vpe2000f2 is for simplex error computing (scalar)

            p;
            start;
            fstring = 'vpe2000mat(fixedparavector,p,nrfittings)';

      %*************************************************************
   ymatfitted = eval(fstring);
      if imag(ymatfitted)~=0
            ymatfitted = abs(ymatfitted)
      end

      % GOFI (absolute error)
      disp('Absolute Error over all points:');
```

```
    err = sum(sum(abs(ymatemp-ymatfitted)))
    disp('Mean Error per fixation point:');
    meanerr= err/(fixedparas(1,1)*nrfittings)

    %computing RMSD

    ymatdiff=ymatemp-ymatfitted
  disp('root mean square deviation');
  ydiff=ymatdiff(:)';
    RMSD= sqrt((ydiff*ydiff')/(fixedparas(1,1)*nrfittings))
  RMSDmat=RMSD*ones(nrfittings,1);
  pmat=ones(nrfittings,1)*p;
  %r-correction
  wl=fixedparas(:,2);
  pmat(:,4)=pmat(:,4)+0.33*(wl-5*ones(nrfittings,1))

  F=[ymatfitted,RMSDmat,pmat];


else
    disp('something is wrong with the length (=nfp) of your vector. Check it out, honey');
end
```

Appendix J3: MatLab-Programs for the NAMWR used for the simplex fitting. The program vpemat2000 used to finally calculate the fitted VPE matrix is identical, but gives back the fitted matrix instead of a deviation value used by Simplex.

```
-------------------------------
function [F] = vpe2000matf2(p,fixedparavector,yvector)

% Computes the VP-Curve as a function of fixation dependent on four
% parameters and four function arguments. The function arguments
%  are set dependent on stimulus
% presentation, the 4 parameters are free and used to fit
% a theoretical curve to the data.
%
% The function arguments are:
% x(1)=nfp: number of fixation positions (default is 5)
% x(2) wl: word length
% x(3)= freq: frequency
% x(4:10) =ngrad

% The 4 free parameters are:
% p(1)=d: drop-off rate: The rate with which recogntion probability drops off:
%    influences height and steepness of the curve
% p(2)=a: asymmetry ratio: The higher "a" the more asyemmtric is the curve.
%   influences the asymmetry of the optimal viewing position
% p(3)=l lexical influence: at any given fixation probability l(1-p(f))*logF
% p(4)=r recogntion at fixation (z-value)


% The function gives a curve for all nfp fixation positions from
% position 1 to nfp


% hcn 3/2000

d=p(1);
a=p(2);
l=p(3);
r=p(4);

%Matrix reshape only for equal nfps yet, else different matrices

nfpequal=fixedparavector(1);

y=reshape(yvector,round(length(yvector)/nfpequal), nfpequal);
nrfittings=size(y,1);

fixedparalength=length(fixedparavector)/nrfittings;
fixedparamatrix= reshape(fixedparavector,fixedparalength,nrfittings)';

nfp=fixedparamatrix(:,1);
wl=fixedparamatrix(:,2);
freq=fixedparamatrix(:,3);
%r=fixedparamatrix(:,4);
ngrad=fixedparamatrix(:,4:(3+ max(wl)));

diff=zeros(nrfittings,nfp(1));
d=abs(d);
a=abs(a);
l=abs(l);
```

```
% creating recognition matrix
for k = 1:nrfittings
   clear R
   for f=1:nfp(k) %fixation position
          for j=1:wl(k) % word length
                %r and a normalized for word length /standard 5
      R(f,j)= normcdf(r+0.33*(wl(k)-5))* (j == ((wl(k)/nfp(k))*f + 0.5*(1- (wl(k)/nfp(k)))))...
          +    normcdf((r+0.33*(wl(k)-5))- (j-((wl(k)/nfp(k))*f + 0.5*(1- (wl(k)/nfp(k)))))*d)*(j > ((wl(k)/nfp(k))*f
+ 0.5*(1- (wl(k)/nfp(k)))))...
          +  normcdf((r+0.33*(wl(k)-5)) - (((wl(k)/nfp(k))*f + 0.5*(1- (wl(k)/nfp(k))))-j)*a*(1 +0.5*(wl(k)-
5)/5)*d)*(j < ((wl(k)/nfp(k))*f + 0.5*(1- (wl(k)/nfp(k)))));
      end
      % lexical and orthographic-lexical correction
      R(f,:)= abs(R(f,:))+(1-R(f,:))*l*wl(k)*(normcdf(freq(k))-
0.5)./(0.5+(normcdf(ngrad(k,1:wl(k))))/mean(ngrad(k,1:wl(k))));

   end


      R=R'; %for the command prod probabilities have to be in the same column
   diff(k,:) = prod(R)-yvector(((k-1)*nfp+1 ):(k*nfp(1)));
       end



% computing RMSD and output function value


diffvector=diff(:)';
RMSD = sqrt((diffvector*diffvector')/(nfp(1)*nrfittings))

F = sqrt((diffvector*diffvector')/(nfp(1)*nrfittings));
```

Appendix J4: MatLab-Programs for the NAMWR: Graph Program. Note that this graph depends on the screen graphic resolution of a given computer.

```
-----------------------------
% makegraphs for different VPE curves

% Define window

% CHANGE position here, in case the window is not defined on your computer screen
figvpe=figure('name','VPE-Fitting Graphs','position', [10 10 1000 690],'resize','on', ...
                                                'units','pixels','numbertitle','off','color','b');

% Number of Figure and legend
Nroffigures=size(posmatrix,1);
legendstr= str2mat('Data','Model');

% Paint single figures
for k = 1:Nroffigures

   %
   wl=ref(k,6);
   RMSD=datasimmatrix(k,6);
   dropoff=datasimmatrix(k,7);
   Asymmetry=datasimmatrix(k,8);
   lexical=datasimmatrix(k,9);
   Recatfix(k)=normcdf(datasimmatrix(k,10));

   %define frequency code for title
   switch ref(k,9)
      case 1
         freqstr='low ';
      case 2
         freqstr='mid ';
      case 3
         freqstr='high ';
      otherwise
         freqstr='???'
   end
   %define ngrad code for title
   switch ref(k,10)
      case 1
         ngradstr='pos ';
      case -1
         ngradstr='neg ';
      case 0
          ngradstr='null ';
      otherwise
         ngradstr='???'
   end

   %define Titlestring

   Titlestr = sprintf('VPE-Fits: %dL, %sFreq, %sNgrad \nRMSD:%.3f, drop:%.3f, Asym:%.2f\n lexical: %.2f
Recatfix: %.2f',...
      wl, freqstr, ngradstr, RMSD, dropoff, Asymmetry, abs(lexical),Recatfix(k));

      ax(k)= subplot(2,Nroffigures/2,k);

      plot(1:nfp,posmatrix(k,1:nfp),'g:+', 1:nfp,datasimmatrix(k,1:nfp),'r-*');
      xlabel('Viewing Position','color','k'); %has to go because of Matlabs xlabel-titleoverlap
      ylabel('Accuracy','color','k');
```

```
  title(Titlestr, 'color','k');
  legend(legendstr);
  %grid on;
  set(ax(k),'xlim',[0 nfp(1)+1],'ylim',[0 1],'xcolor','k','ycolor','k');
  set(gca,'YGrid','on', 'xtick',[1 2 3 4 5 ])
end
```

Appendix K: Reference matrix used for model fits over all experiments.

Exp.	WF	VP1	VP2	VP3	VP4	VP5	WL	WFC	SC	FC	NGC	logF	Ngrad	Syl.	Mask	N	Npos1	Npos2	Npos3	Npos4	Npos5	Npos6	Npos7
PFH	ALC	0.51	0.62	0.67	0.55	0.44	5	1	1.00	3	0	2.02	-0.24	1.7	7	2.68	1.38	0.38	0.38	0.34	0.20	0.00	0.00
PFH	IUC	0.69	0.76	0.79	0.71	0.54	5	2	1.00	3	0	2.02	-0.24	1.7	7	2.68	1.38	0.38	0.38	0.34	0.20	0.00	0.00
PFH	AUC	0.55	0.71	0.66	0.65	0.45	5	3	1.00	3	0	2.02	-0.24	1.7	7	2.68	1.38	0.38	0.38	0.34	0.20	0.00	0.00
PFH	ALC	0.70	0.85	0.81	0.72	0.55	5	1	0.00	3	0	1.75	-0.06	1.68	7	1.86	0.60	0.18	0.60	0.14	0.34	0.00	0.00
PFH	IUC	0.75	0.83	0.75	0.72	0.57	5	2	0.00	3	0	1.75	-0.06	1.68	7	1.86	0.60	0.18	0.60	0.14	0.34	0.00	0.00
PFH	AUC	0.55	0.73	0.79	0.67	0.51	5	3	0.00	3	0	1.75	-0.06	1.68	7	1.86	0.60	0.18	0.60	0.14	0.34	0.00	0.00
WLF	AUC	0.49	0.59	0.64	0.59	0.44	3	3	0.80	1	0	0.18	-0.16	1.20	9	2.76	1.44	0.20	1.12	0.00	0.00	0.00	0.00
WLF	AUC	0.57	0.75	0.71	0.75	0.59	3	3	0.26	3	0	2.76	-0.28	1.00	9	2.72	1.44	0.40	0.88	0.00	0.00	0.00	0.00
WLF	AUC	0.49	0.64	0.65	0.43	0.16	5	3	0.92	1	0	0.18	-0.27	1.76	9	2.80	1.60	0.36	0.36	0.12	0.36	0.00	0.00
WLF	AUC	0.63	0.79	0.80	0.72	0.37	5	3	0.28	3	0	2.86	-0.05	1.60	9	2.72	0.76	0.20	0.84	0.60	0.32	0.00	0.00
WLF	AUC	0.16	0.52	0.52	0.20	0.00	7	3	0.44	1	0	0.18	-0.05	1.92	9	2.72	0.56	0.20	0.44	0.64	0.32	0.08	0.20
WLF	AUC	0.53	0.79	0.75	0.52	0.25	7	3	0.30	3	0	2.50	-0.06	2.00	9	2.72	0.80	0.04	0.72	0.36	0.36	0.16	0.28
N-Grad	AUC	0.56	0.59	0.65	0.51	0.32	5	3	0.76	1	-1	0.42	-1.19	1.8	7	7.72	5.72	1.04	0.64	0.04	0.28	0.00	0.00
N-Grad	AUC	0.75	0.76	0.81	0.73	0.56	5	3	0.84	2	-1	1.23	-0.98	1.96	7	6.92	4.64	1.00	0.88	0.28	0.12	0.00	0.00
N-Grad	AUC	0.77	0.81	0.84	0.71	0.65	5	3	0.56	3	-1	2.30	-1.08	1.76	7	7.84	5.24	1.12	1.04	0.12	0.32	0.00	0.00
N-Grad	AUC	0.36	0.56	0.57	0.52	0.28	5	3	0.92	1	1	0.38	0.48	1.56	7	3.00	0.00	0.04	0.12	0.84	2.00	0.00	0.00
N-Grad	AUC	0.51	0.67	0.72	0.64	0.44	5	3	0.80	2	1	1.26	0.42	1.44	7	3.52	0.20	0.00	0.64	0.76	1.92	0.00	0.00
N-Grad	AUC	0.53	0.72	0.80	0.81	0.41	5	3	0.52	3	1	2.24	0.30	1.48	7	2.68	0.04	0.12	0.40	1.00	1.12	0.00	0.00

Exp.: Experiment, WF: Word form, VP: Viewing position, WL: Word length, WFC: Word form code, SC: Syntactic Class (Nouns=1, Non-Nouns= 0), FC: Frequency Code, NGC: N-Grad Code: log: Logarithmic frequency per million, Syl.: Number of syllables, Mask: Number of mask #, N: Number of Neighbors, Npos<i>: Number of neighbors on position i (only the neighbors on positions i=1,..., word length were read out). Used for the fittings were word length, log, N and the positional neighbors (see formula). Other stimulus attributes may need to be considered in future versions of the model.

Appendix L: Parameters used in the models of this thesis:

Parameter	Status	Equations	Description
A	free	1, 3-7, 9-14	Asymmetry in the drop-off rates to the left and right (d_{left} / d_{right})
d	free	1, 3-14	Drop-off rate to the right from fixation
f	fixed	1-2, 7-8	Indexes the fixation position on which the word is presented
Freq	fixed	10-11, 13-14	Total word frequency per million
i	fixed	3-14	Indexes the (ordinal) letter position (right from spatial fixation position in Equations 6-8)
j	fixed	6-8	Indexes the (ordinal) letter position (left from spatial fixation position in Equations 6-8)
k	fixed	1,8	Indexes the letter distance between the fixated letter and a letter to be recognized
l	free	10-11,13-14	Lexical parameter being adjusted to the data
$\overline{N}$	fixed	11	Mean number of neighbors per letter position
n_f	fixed	2, 7-8	Number of fixation positions (= word length in Equation 1 only).
posN	fixed	11,13-14	Vector of length w whose values are the (average) number of neighbors on each letter position
r_f	fixed	1, 3-8	Letter recognition probability on fixation: Fixed to 1
r_f	fixed/free	9-14	Normally distributed z-value for the letter recognition probability on fixation: Fixed in Equations 9-11, free in the final model Equations 12-14.
w	fixed	1-2, 5-8, 10-14	Word length
y	fixed	2-6, 9-14	Spatial letter (fixation) position in a word whose space is ranging across spatial letter positions [0.5, w + 0.5]

All parameters in alphabetical orders, their status (free or fixed), the equations in which they appear, and a short description of their function are depicted in this appendix. Note, however, that some parameters are a function of others (e.g., y is a function of f, w, n_f), and could be inserted into equations in which they do not actually appear.

Acknowledgements

I am truly indebted to many people who inspired or suffered with me at some time in the dissertation process. Without their help and support, this thesis would never have been written.

Thanks are due:

- First and foremost to my wonderful Dissertation Support Group (DSG). Kerstin, Martina, and Suse helped to set realistic goals and to continue at a time when the end of the tunnel could not be seen.
- To Prof. Arthur M. Jacobs whose advice and criticism shaped this thesis and whose generous agreement to my change to Aachen made it possible for me to finish this thesis.
- To PD Dr. Siegfried Gauggel who supported and helped me in Aachen and whose invitation to stay for two weeks in Chemnitz for all the major revisions helped to overcome the last major obstacles.
- To Prof. H.-H. Schulze for his help and advice, in particular with regard to the NAMWR.
- To Prof. K. Willmes-von Hinckeldey for his patience and support who gave me time to finish the thesis even though much (more) working time should have been spent in the projects I was appointed for.
- To Daniela Brustmann, Dirk Lehr, Katja Oßwald, and Kai Richter for their help in data acquisition.
- To Emma Stringfellow for proofreading for English language errors.
- To Ralf Graf, Annette Kinder, Tatiana Nazir, Ken Paap, Arnaud Rey, Florian Schmidt-Weigand, and many others not mentioned here for inspiring and appreciated discussions that helped to improve this thesis.

In addition, I want to thank all the friends around me whose friendship and emotional support helped me in the last four years. My gratitude goes in particular to:

- Annegret who made me sleep, love, live and continue.
- Ellen who made me feeling at home in the Lichtenholz in rough times.
- Martin, Ralf, and Regine for four years of continuous support and close friendship in times of hope and worries that certainly were not always easy to handle.
- Alfred, Andre, Anna, Christiane, Christine, Claudia, Daniela, Dominik, Florian, Francine, Heinz, Kathrin, Mario, Marion, Markus, Matthias, Rainer, Ralf, Sabine, Werner, and many others whose help, support, and share of enthusiasm and pain was important for me in some meaningful hours.
- Once more to my wonderful DSG for sharing “the sweet and the bitter” with humor and sympathy. I am looking forward to starting our LSG soon.

statt eines Nachworts

sollen Sie erfahren, welchen Metamorphosen Übersetztes
unterliegen kann.
Im Jahre 1902 wurde Goethes berühmtes „Nachtlied“:
„Über allen Gipfeln
Ist Ruh.
In allen Wipfeln
Spürest Du
Kaum einen Hauch.
Die Vöglein schweigen im Walde.
Warte nur, balde
Ruhest Du auch.“
ins Japanische übertragen. Neun Jahre später übersetzte
ein Franzose in der irrigen Meinung, es handle sich
um ein japanisches Originalgedicht, die Goethe-Verse
aus dem Japanischen ins Französische.
Von einem Nachdichter fernöstlicher Lyrik
wurde das Goethe-Gedicht schließlich
ins Deutsche zurückübersetzt
und als „Japanisches Nachtlied“
in einer deutschen Zeitung veröffentlicht. Aus den
Versen ist mittlerweile folgendes geworden:
„Stille ist im Pavillon
aus Jade.
Krähen fliegen stumm
zu beschneiten Kirschbäumen
im Mondlicht.
Ich sitze und weine.“

(a German poem about the metamorphoses of translations, and translated poems, in particular; from the program to „Danzón“ of Pina Bausch)