

Optimality Theory: rethinking phonology for the study of literacy

Per Henning Uppstad

1 Introduction

In a traditional view, written language is simply considered a mirror of spoken language both in structure and function (Bloomfield 1933). According to this view, spoken language is structured as a linear sequence of more or less clear-cut sound units, with a relatively close correspondence to the alphabet. Furthermore, the overall function of written language is to represent speech. This traditional view is to a certain extent maintained in the scientific study of literacy, although major discoveries in phonology over the past 20 years (Horne 2000) have questioned the linear and segmental approach to spoken language. The fact that linear theory cannot incorporate major findings in phonology highlights the request for an alternative framework for the structure of spoken language.

In this paper I will investigate some possibilities lying in an alternative framework for the structure of spoken language, and therefore for the understanding of the relation between written and spoken language. It is at this point that it becomes relevant for the the study of literacy, and the term literacy is here used in a broad sense. It is an important argument in the study of literacy that phonological structure can be described very differently from traditional linear theories. Bringing alternative phonological theories into the study of literacy may provide new assumptions of the interplay of spoken and written language and may also challenge some existing views. I will stick to findings in modern phonology and present optimality theory as a competing theory for the study of literacy.

The traditional view is common knowledge in many scientific disciplines working with literacy, still it has been challenged. Modern phonology's rejection of a strict linear and segmental structure of spoken language therefore also represents a severe argument against the traditional view of written language. Still, this argument is not a hot issue in linguistics, a fact

that probably is derived from the linguistic tradition's focus on a systematic description of spoken language without regard to the relationship between spoken and written language. As a consequence, there are several descriptions of structure that can be considered valid. The scientific study of literacy probably has a stronger demand for psychological reality of the description than does theoretical phonology. Therefore this tradition should be eager to underline – but not simply adopt – the findings in modern phonology in order to question and adjust the present description of spoken language.

2 Optimality Theory – leaving the segment unfocused

Optimality Theory by Rene Kager gives a good introduction to the establishment of this branch of generative grammar through the 1990s. At the end of the book he shows the theory's major deviation from its origin by stating that Optimality Theory (OT) has kept only one of three main assumptions in generative grammar: "i) all contextual variants of a morpheme derive from a single Underlying Representation", rejecting: "ii) by rewrite rules ($A \rightarrow B / X_Y$), iii) applying in a serial derivation" (Kager 1999:413). In this paper, I will present the OT view on all three assumptions, starting with the rejected ones. The OT solutions to these assumptions form in many ways the OT framework.

Having stated this situation, Kager points out a tendency to head further away from the generative origin in the development of OT. This further deviation questions the notion of Underlying Representation, looking for a solution that might be more in harmony with central OT assumptions. If this solution can be shown adequate, OT has rejected all main assumptions of generative grammar.

Further, I will focus on some aspects of Optimality Theory that may both keep the theory of grammar based on universal elements of language and may bring generative grammar out of its strict axiomatic corner. My intention is to point out parts of optimality theory that may support studies in linguistic areas that claim to be psychologically valid in some way. Such intentions are controversial in theoretical linguistics and therefore the subjects of *language system* and *language use* are traditionally kept apart, mainly through Chomsky's division between competence and performance, but also by a strict division between phonology and phonetics:

"Therefore a model of grammar is adequate to the extent that it explains observed systematicities in natural languages, and grammatical judgements of speakers. Explaining the actual processing of linguistic knowledge by the human mind is not the goal of the formal theory of

grammar, but that of linguistic disciplines. The central point is that a grammatical model should not be equated with its computational implementation" (Kager 1999:26).

Surprisingly, in his otherwise sound defence of OT, Kager suddenly claims 'non-mentality' in a situation where it is not necessary, and where it over-shadows the potentials of OT. Kager is right in some way, but this non-mentalistic attitude blocks the potential use of theoretical grammar in other linguistic disciplines. At the same time, it ought to be problematic for a generative grammar to reject proposals of mental processing and leave Chomsky's axiomatic model unsolved. Therefore, the goal for a formal theory of grammar should primarily be the structure of natural languages. However, the possible relevance for processing should not be rejected, as Linell (1979:8) points out:

"The kind of linguistic theory needed in the description and explanation of language acquisition, foreign language learning, speech performance (including, e.g., speech errors and aphasia) must of course be psychologically valid"

There are plausible reasons to look upon *competence* and *performance* in a more connected way than has been done in traditional generative grammar. In many ways, the strict division between *competence* and *performance* is contradictory to what I consider the potential in OT. As a theory of constraints evaluating competing surface forms, OT competence is highly oriented towards performance. This view is supported by the OT notion of violable constraints – in contrast to a ruled-based generative grammar. By the fact that universal constraints can be violated, OT can to some extent explain why some sub-optimal forms occur. In this way, OT grammar embraces an aspect of variation on a scale, where traditional generative grammar pushes sub-optimal forms into the *performance* sphere – and therefore out of grammar. As a consequence, a great part of systematic language is placed outside grammar in rule-based theory. Still, every theory of language needs a distinction of acceptable and non-acceptable forms. If the OT explanation of processes is right, OT theory is capable of embracing a larger part of language in the concept of grammar than can be done in traditional generative grammar. This fact forces a looser distinction between *performance* and *competence*, or – in other words – it links up grammar with its implementation, contradictory to what Kager claims initially.

2.1 OT framework and the rejected assumptions

As stated above, the OT solutions to the two rejected assumptions form the main framework of OT. To see this, we need to understand the overall perspective of OT and how this contrasts with traditional generative grammar. OT grammar claims that natural languages consist of universal constraints, where the ranking of constraints is language-specific. Therefore, learning a certain language means learning the specific ranking of constraints. OT considers language as a *system of conflicting universal forces*, roughly presented as *markedness* and *faithfulness*. An OT grammar is essentially an input-output mapping device (Kager 1999:404) where faithfulness is a force preserving lexical contrast, and markedness is a force striving at naturalness. The latter is grounded on articulatory and perceptual facts, in which a segment can be either marked or unmarked. Every language demands a certain level of lexical contrast, and is able to sustain a certain level of markedness – or ‘unnatural’ features. In other words, every language will be redundant and unnatural to some extent. The two forces are the forces that balance functionality on the one hand, and perceptual and articulatory ability on the other. Kager clearly shows this by two hypothetical cases, one giving maximal priority to faithfulness, resulting in a totally unnatural language with an enormous amount of potential lexical items, the other giving maximal priority to markedness (keeping items unmarked), resulting in an acute shortage of lexical contrast. Both are implausible in different ways.

“A language can be maximally faithful to meaningful sound contrasts only at the expense of an enormous increase in phonological markedness. Conversely, a language can decrease phonological markedness only at the expense of giving up valuable means to express lexical contrast” (Kager 1999:6).

These forces are described as families of constraints working in the two directions. Hence, we have a large amount of faithfulness- and markedness constraints. Constraints only work on output forms, produced by an assumed function called *Generator* (GEN). GEN produces multiple candidates for evaluation on the base of the input, only limited by the human speech apparatus. This is a ‘wild’ function, noted as ‘Freedom of analysis’ in OT.

“Since Gen generates all logically possible candidate analyses of a given input, the OT grammar needs no rewrite rules to map input onto outputs. All structural changes are applied in one step, in parallel. The evaluation

of these candidate analyses is the function of the Evaluator, the component of ranked constraints discussed in 1.4.3” (Kager 1999:20).

In OT constraints are violable, which means that even optimal forms can violate some constraints. The optimal form is the form that has less violation marks than the other candidates. Due to the constraint hierarchy, an optimal form may even have a larger amount of violation marks than other forms and still be optimal, because the sub-optimal forms violate a higher ranked constraint than the optimal form.

The notion of *Evaluator* (EVAL) embraces both the hierarchy of constraints and the devices for evaluation. These devices use the hierarchy of constraints; first violation marks are assigned on outputs for every constraint, and secondly all the outputs are ranked in respect to harmony, and we get a scale which span from the less harmonic form to the optimal (most harmonic) form. In this discourse, EVAL is presented as a concrete thing, as some kind of mechanical device. Still, we are not able to falsify whether there is such a thing inside our minds. Nevertheless, the notion of *evaluator* may function as a metaphor for crucial evaluating processes in language production and perception. And more importantly – the OT notion of EVAL may bring us closer to a plausible explanation than traditional generative theory because EVAL is not introspective, rather, it operates on observable output forms. In other words, it is the function of the constraint hierarchy.

As quoted above, the function of the GEN rules out rewrite rules in generative grammar. In addition, this whole new perspective has no need for building output forms from underlying forms. On the contrary, constraints work on output forms produced by GEN. In this way, OT rejects the generative assumptions of rewrite rules. The serial aspect (3rd assumption) is rejected at the same time, because there is a *direct mapping* between input and output, directed by constraints. Still, there are issues in grammar that can be dealt with by serial analysis, e.g. opacity, which have not yet been explained adequately in OT. The rejection of rewrite rules is also based on the ‘output blindness’ of rules in generative grammar. Rules are not surface sensitive. On the contrary, constraints work on output/surface forms only:

“This predicts that no property of phonological forms depends on information that is not present in the output – either in the output alone, or in the relation between the input and the output” (Kager 1999:58).

2.2 Underlying representations – the ‘remaining’ assumption

By this coarse presentation of the framework of OT, the rejection of the two generative assumptions should be clear. Turning to the remaining assumption – the one of underlying representations – Kager claims introductory to his book that “Optimality theory is not a theory of representations, but a theory of interactions of grammatical principles. More accurately, the issue of representations is orthogonal to that of constraint interaction” (Kager 1999). Nevertheless any status of the underlying or input representation diverging from traditional generative grammar is an interesting deviation from generative theory. In fact, Kager’s notion of the orthogonal relationship between representations and constraint interactions, clarifies a point in OT: the input representation exists in another way than in traditional generative grammar. The OT input representations exist as possible output forms. In the original theory the representation was *constructed* even beyond possible human language forms, in order to fit claims of a simple and consistent theory.

By the fact that it has been constructed, the underlying representation (UR) in traditional generative grammar has the character of a main argument in a serial reasoning. This is evident in the cases where the UR is abstracted to contain features which actually never surface in the current language, like the segment /x/ in /nixtVngæl/ (Kaisse & Shaw 1985:15). In OT the role of UR is different, even if Kager claims that this is the only remaining assumption from traditional generative grammar. Early OT does not seem to have any explicit deviation from the original, but as a tendency the UR is closer to the surface than its origin. This is not at all surprising due to the output focus in OT, and the recent OT questioning of the UR can be considered as a reasonable consequence of the new framework. The notion of UR in early OT does not have a serial character, because it is not the main argument in some serial reasoning, a function which in OT is replaced mainly by GEN and EVAL. This might clarify the orthogonal relationship between representations and constraint interaction. In this perspective, the OT view of representation is important – because there is no representation being constructed: it’s simply there. Still, the shape of input forms is not random. The learning algorithm of Tesar and Smolensky shows the importance of selecting the right underlying representation – or input, a central idea in the notion of *Lexicon Optimization*.

The recent questioning of the role of UR, suggests removing the underlying representation, replacing it with output forms in a theory of

allomorphy. In this new orientation, one no longer counts with a unique underlying representation, but with allomorphs mapping from one surface form to another. The notion of Lexicon Optimization can be considered as a goal in OT, being an input-output mapping device: the mapping is easy if output is equal to input.

“However, the key difference from the standard model resides in a new role of the lexicon. The lexicon no longer supplies a unique UR for each morpheme, but instead it supplies a set of shape variants of the morpheme, allomorphs, chunks ready for insertion in various morphological contexts (base or affixed forms)” (Kager 1999:415).

This means that each allomorph has its lexical entry. If this new role of the lexicon can be shown adequate, it represents an important upgrade of surface forms as symbols, at the cost of innate and ‘hidden’ elements. In this perspective, the lexicon is also clearly constituted by relation to other output forms, or in metaphorical terms: the lexicon is by its surface-sensitivity more ‘active’ – not like a hidden treasure. In this new orientation, the phenomena of alternating morphemes has changed explanatory area from viewing inputs as (more or less) abstract forms to observable output forms. As a consequence, the alternating morphemes have to be explained with a stronger focus on output-to-output (OO) constraints. Due to the orthogonal relationship between representation and constraint interaction, the role of representation/input can be easily revised without fatal consequences – a fact supporting OT as non-serial. Kager also states that this displacement of explanation changes the role of the grammar as an input-output mapping mechanism into a checking mechanism (Kager 1999:414):

“In this UR-less model, the input (lexical shape) simply equals the output (surface shape). IO-faithfulness maintains its original function of reinforcing parts of the input, protecting it against the neutralizing forces of markedness”

As the UR has been removed, morphological related forms like (Dutch) *bet* and *beden* are not connected in a phonological sense, and this is where OO-correspondence becomes active.

“Two sets of faithfulness constraints are assumed in this model. OO-correspondence is active in its standard role of checking identity in the network of morphologically related output forms, or the paradigm. IO-faithfulness checks identity between allomorphs in the lexical input and their output counterparts” (Kager 1999:415)

Due to the lack of a shared underlying representation, output candidates have to be evaluated in paradigms or in relation to each other. OO-correspondence works by evaluating both base and affixed form simultaneously. The two faithfulness constraints have different tasks, as shown in the tableau below.

Input: {bed ~ bed-en}	*VOICED CODA	IDENT-IO (VOICE)	IDENT-OO (VOICE)
a) \varnothing bet ~ bed-en		*	*
b) bet ~ bet-en		**!	
c) bed ~ bed-en	*!		
d) bed ~bet-en	*!	*	*

*VOICED CODA = constraint that assigns violation mark if coda is voiced

*IDENT-IO (VOICE)= constraint that assigns violation mark if input and output differs with regard to voiced coda

*IDENT-OO (VOICE)= constraint that assigns violation marks if the two output forms differ with regard to voiced coda

* = violation mark

! = fatal violation

We assume that both GEN and EVAL work on a mini-paradigm, here represented by base form and the plural form, and that the OO-correspondence functions in this way. The constraints therefore work both on each morpheme (IDENT-IO) and between related output forms (IDENT-OO) as a checking device. More precisely, in the candidate paradigm a) IO constraint checks the output *bet* with the input *bed* and the output *bed-* with the input *bed-*, and the OO-constraints check the two morphologically related output forms *bet* and *beden*. As we can see in the optimal form, both IO and OO constraints are violated, but this paradigm is still the candidate pair with fewer violation marks.

In Kager's view the UR-less model offers two major advantages: primarily a reduction of abstractness of lexical representation – and therefore an increase in cognitive plausibility – and second a uniform analysis of all types of alternations. The latter is a result of the new role of the lexicon in this model.

“The lexicon is the place where idiosyncratic properties of a language are listed, so nothing guarantees that every morpheme will indeed exhibit the

full range of allomorphs as compared to other morphemes” (Kager 1999:419).

Still, he states clearly that this model cannot capture drastic alternations like for instance *go – went*, which might open for a less strict model with stratified analysis. The solution of stratified analysis seems to be an iterative one, since it occurs in many of Kager's residual issues. The different OT-solutions to opacity, free variation and positional faithfulness can be seen as attempts of loosening up a somewhat strict deterministic theory by upgrading the lexical influence and the interaction between output shapes. Probably, these issues remain challenges to the theory even in a UR-less model. As an example of remaining challenges, let us turn to the issue of variation:

“Another unresolved issue is variation (...) By definition, the distribution of both outputs cannot be under grammatical control, since that would restore (ranking based) determinism in the choice of both variants” (Kager 1999:404).

As a residual issue Kager discusses how to deal with free variation in grammar. As he admits, this problem is neither unique to OT nor to derivational grammar or any theory of grammar (Kager 1999:405). In fact, variation offers a challenge to every theory that claims to be well formed and deterministic. A consequence of strict determinism in grammars is divisions like *langue/parole* and *competence/performance*. Dealing with grammar or systems, one cannot simply overlook such demarcations. But still, one may discuss whether the demarcation is too strict, or whether the consideration of a well-formed grammar should overrule the description of systems in a wider range. It probably is possible to loosen up the division of competence and performance and still keep the ambition of a deterministic grammar. Kager asks, “how can free variation be reconciled with the deterministic nature of the grammar?” (Kager 1999:404). He presents two different suggestions: *co-phonologies* and *free ranking*. The idea of co-phonologies has its origin in lexical phonology, and has therefore a serial character. In co-phonology we have a common input to independent, parallel constraint hierarchies that may produce radically different outputs. Kager claims that this is incorrect for free variation because outputs in free variation mostly vary minimally.

The second suggestion, *free ranking*, seems surprisingly equal to the first one, by proposing “Evaluation of the candidate set is split into two subhierarchies, each of which selects an optimal output. One subhierarchy has $C1 \gg C2$, and the other $C2 \gg C1$ ” (Kager 1999:406). Still, these

hierarchies differ from the former by having only some free ranked constraints, and are still governed by 'hard' constraints in grammar. This limitation may explain the observed similarities in variable outputs. The hierarchies may also be quite different, selecting quite different outputs. But when they differ, the solution of free ranking supplies a correlation between the degree of dissimilarity and the number of variable outputs (Kager 1999:407). The prediction of this correlation is not possible in co-phonology, because those hierarchies are independent from one another, while in *free ranking* a certain number of free-ranked constraints give a predicted number of optimal forms. The prediction of variable forms may satisfy the demands for a somewhat deterministic grammar. Kager presents the latter solution as a promising one, but raises the empirical question whether this kind of grammar is learnable. Kager states, that if possible, it will require major adaptations of Tesar and Smolensky's learning algorithm (Kager 1999:407).

"The fact that variation is 'free' does not imply that it is totally unpredictable, but only that no grammatical principles govern the distribution of variants. Nevertheless, a wide range of extragrammatical factors may affect the choice of one variant over the other, including sociolinguistic variables and performance variables. Perhaps the most important diagnostic of extragrammatical variables is that they affect the choice of occurrence of one output form over another in a stochastic way, rather than deterministically"

As a future perspective, Kager points out the tendency for functional explanations, a direction that forces a more intertwined conception of performance and competence. In this context, he presents the ongoing blurring of the phonology-phonetics boundary. This can be seen in connection with the surface-focus in OT, that is: constraints on features in words – in many ways a more phonetic theory than former generative and structuralist theories. Even if there are major problems in giving up the distinction between phonology and phonetics, OT shows that the strict distinction can be challenged. The developing line scheduled here – removing the UR and blurring the distinction between phonology and phonetics – brings forth demands for a new explanation of symmetry in the phonological system (Kager 1999:421). A heavy focus on perceptual and articulatory factors in grammar, forces a solution that may also capture their gradient character, a direction opening for more connectionist explanations:

"When numerical values (of phonetic parameters) are incorporated into phonology, it seems only logical to take another step into numerical (gradient) types of interaction, so that constraints exert influence

according to their relative strength, expressed in an index" (Kager 1999:422).

Similar approaches can be found in studies of language acquisition. In a paper of Lacerda and Lindblom, it is claimed that variability should not be looked upon as noise, because variability – for whatever reason – is systematic. Therefore variability becomes the primary source of information.

"According to the model, categories with prototypical properties emerge in a self-organizing way as a consequence of the distributional patterns that the exemplars cumulatively form in the perceptual representational space and in phonetic memory" (Lacerda & Lindblom 1997:14).

If we do agree with the importance of variability for building a system, we should dare to hypothesize the importance of variability in going the opposite way, in dealing with the system's relevance for perception and production. If not, we risk assuming that linguistic input is of quite another quality than linguistic output. Turning away from the understanding of variability as noise, we move towards a less abstract and more specified explanation, a potential insight in Lacerda and Lindblom's conception of speech sound as *fuzzy clouds*:

"It is the implication that categories of speech sounds are fuzzy clouds of multidimensional (multimodal) sensory representations that are spontaneously structured by cross-correlation, due to their similarity" (Lacerda & Lindblom 1997:31).

All these contributions clearly uncover the assumption that the language learner does not have any linguistic knowledge at time of birth – instead language is built from the variability in the linguistic and non-linguistic surroundings. The learning-algorithm of Tesar and Smolensky is built on the same assumption, even if it – being a highly deterministic device – still cannot adequately predict forms of variation. In OT, the notion of violable constraints is a central part of the theoretical foundation, which means that even optimal forms have violation marks. Therefore, OT is a gradient description of forms. Violation of constraints is also a cornerstone in the learning algorithm, even if it assumes that the input to the learner has to be consistent. The problem is, that human language input is never consistent in this sense. An interesting thought is whether probability or more numerical approaches can complement the algorithm.

In Kager's presentation of the algorithm, he makes an effort to show how much information about constraint ranking is present in a single optimal

form. As an example this focus is convincing, but in the context of language learning it is quite bizarre to focus on learning effect from one single word. In his presentation, any variability is also abstracted away. This is probably necessary to make the algorithm work, but one could be tempted to add varying forms to this picture, assuming that surfacing sub-optimal forms also contain important information about constraint violation. Not only in the sense of constraints being violated alone, but accompanied by information about probability of occurrence.

This can be seen in connection with what Lacerda and Lindblom 1997 formulate as consequences of phonetic memory. The learning algorithm is an algorithm of upgrading and degrading of constraints, in which the violation of a constraint gives far more important information than constraints that are satisfied. In fact, it is the violated constraints in an optimal form that constitute the grammar in the learner. This is simply because violation plots constraints into the hierarchy, while satisfaction conceals. The learning algorithm seems – in its simplified version – to be a plausible explanation of how grammar can be learnable from output forms. And further, this approach offers a solution different from traditional generative grammar in the area of universality, directing us away from the innate and therefore strictly determined path of language. Kager also signalises this:

“The learnability of constraint ranking is a crucial ingredient of the explanation of learnability of grammars. It is only an ingredient, however, since much more is at stake in learning a grammar than constraint ranking” (Kager 1999:322)

Therefore, the algorithm should only serve as a model for how constraint ranking proceeds, and the importance of its incapability of describing variation should not be overestimated:

“We should however, acknowledge the danger in this: that researchers put too deeply into models that are far too simple. We must always remember that the point of using any heuristic tool in science is ultimately to throw it away” (Tønnessen 2002:94)

2.3 Word forms as primes

Due to focus on output representations, OT contributes to a revitalized understanding of spoken language: *word forms as primes*. This aspect may not be intended in the theory, but emerges as an effect that in many ways manages to keep phonology functional. One central idea in OT is the *direct mapping* from input to output, which means that there are no serial aspects like rules

and intermediate levels. The direct mapping underlines the work of EVAL, which is to evaluate output forms on the basis of constraints. EVAL – the hierarchy of constraints – is therefore a licenser of optimal word forms. In this thinking every phonological property is brought in connection with the word as a whole. Further, the generative notion of underspecification has come under heavy pressure in recent OT: the constraints work on phonetic features in the output candidates. This implies a role of the phoneme that differs from cognitive science, which holds that phonemes are stored as mental representations. In OT the traditional conception of the phoneme – if used – is only relevant in the context of words or morphs, and clearly subordinate to the hierarchy of constraints. Inherent in the conception of the phoneme is a strict division of phonology and phonetics, a division that is being blurred in recent OT. Due to this blurring, the OT understanding of the acoustic signal may be considered as a whole, *consisting of* a number of segments and a certain number of features, while in generative grammar segments or phonemes with their features *form* a word. This distinction is an aspect of the contrast between the serial aspect in generative grammar and direct mapping in OT. Therefore, the direct linking to autonomous segments (not features) is no longer in focus: constraints count. It is difficult to consider this understanding of the phoneme as ‘new’, because it can easily be compared to pre-generative theory. Saussure's understanding of the phoneme is in many ways problematic – as mentioned above – but still, his phoneme had a value only as part of a word, contrasting other words in minimal pairs. Saussure had an output focus on functional forms and output contrast, but the theory ended up in an abstract and very segmental model.

“By way of summary, I would say that the segmental analysis, which is nearly always taken for granted in phonology, is not self-evident. Segments may be perceived as units, but such an analysis is probably partly a result of cultural traditions (alphabetical writing, the tradition of talking about sound structure in terms of ‘sound types’). There are several facts that speak for the perception of morphs and word forms as phonetic gestalts. In fact this is precisely what may be expected, given that (morphs and) word forms are communicative and grammatical primes” (Linell 1979)

In Linell's terms, *segmental* probably overlaps the notion *linear* and even *serial*. His work appeared before the occurrence of Autosegmental–Metrical theory (AM), which claims a non-linear phonology, but keeps aspects of segmental analysis. This theory was able to split the notion of segmental analysis and linear analysis, accepting the first and rejecting the latter. The

so-called *feature geometry* is a position inside AM-theory that seizes the segment but claims auto-segmental association of features (Gussenhoven & Jacobs 1998, Roca & Johnson 1999). In this position features are hierarchically organized, and may be considered as constraints of association. The crucial point here is that the segment in feature geometry is focused, while in OT it is not. Although the notion of the segment is not denied in OT, one could say that in OT, it is output variation that is focused on by the theory. Thus, preferences in choosing position in phonological theory will be guided by preferences of the general theory. In my view, the OT position has a better potential for the connectionist notion of lexicon. Still, we cannot simply reject the importance of the segment. What we can do is to question whether the segmental analysis has been overestimated in generative phonology. Several researchers have pointed out such tendencies (Öhman 1979, Linell 1979, Liberman 1999). Liberman claims that phonetic representations are primary percepts, and Linell claims that that psychological reality in phonology is in the shape of so called 'phonetic plans', defined as 'the most careful pronunciation'. Still, Linell does not reject the importance of segments and phonemes.

The main issue is probably not a question of one or the other view, but a question of domination in which both appear. Therefore, the important question could be posed as: is production and perception of spoken language primarily serial/linear, or primarily direct mapping of output forms? OT theory claims the latter, in opposition to the generative tradition. This is a theoretical change of course, a course which is probably not groundbreaking for an isolated description of language, but which may have great importance for what Kager calls linguistic disciplines. This change of direction is towards more functional aspects of language, due to the theory's focus on output forms.

3 Conclusion

OT offers a description of language that is based on observable forms – not on hidden principles and parameters. Concerning assessment of language and literacy skills, this point becomes very important in the aspect of reducing the number of assumptions connected to structure, and thereby avoiding introspection. Together with the increase of cognitive plausibility the theory might have a potential in being a common framework for both linguistic and psychological approaches to spoken language. In optimality theory, and also

by Linell, it is claimed isomorphy between the internal grammar and representations – there are no underlying non-altering morphemes.

The resemblance between OT and structuralism is mentioned by Kager, in the view of allophonic patterns, contrast, surface patterns, functional considerations, and allomorphy (Kager 1999:422). Several places I have pointed at common insights between OT and Linell's criticism of orthodox generative grammar, and he himself also points at the apparent similarity between his point of view and structuralism – as well as he clearly states the differences. Still, history is not iterative, even if some important insights have been lost on the way, and later been rediscovered. In this gap of time, we have come to know more about the world's languages. Also, some major discoveries have been made, which opens for a better description and explanation of language. The relevance of the theory for the study of literacy lies in providing a theory of spoken language that comprises the non-segmental part of language structure. Therefore, the relevance of a new framework for the study of literacy is to underscore the different character of spoken and written language. With the convenience of segmental theory for the alphabet in mind, this endeavour might seem reactionary both related to pedagogical issues and to the scientific principle of Occam's razor. Still, the demand for consistency and the challenge from new findings are good reasons to take a step back before we can take further steps forward.

References

- Bloomfield, Leonard. 1933. *Language*. New York: Holt.
- Gussenhoven, Carlos & Haike Jacobs. 1998. *Understanding phonology*. London: Arnold.
- Horne, Merle (ed). 2000. *Prosody – theory and experiment: studies presented to Gösta Bruce*. Dordrecht: Kluwer.
- Kager, René. 1999. *Optimality theory*. Cambridge: Cambridge University Press.
- Kaisse, Ellen M. & Patricia A. Shaw. 1985. 'On the theory of lexical phonology'. *Phonology yearbook* 2, 1-30.
- Lacerda, Francisco & Björn Lindblom. 1997. 'Modelling the early stages of language acquisition'. In Å. Olofsson & S. Strömquist (eds), *Cost A8 Cross-linguistic studies of dyslexia and early language development*, 14-34. European Commission.
- Liberman, Alvin M. 1999. 'The reading researcher and the reading teacher need the right theory of speech'. *Scientific studies of reading* 3, 95-113.
- Linell, Per. 1979. *Psychological reality in phonology: a theoretical study*. Cambridge: Cambridge University Press.

- Öhman, Sven. 1979. 'Introduction: Gunnar Fant, pioneer student of human speech'. In S. Öhman & B. Lindblom (eds), *Frontiers of speech communication research*, xv-xxv. London: Academic Press.
- Roca, Iggy & Wyn Johnson. 1999. *A course in phonology*. Oxford: Blackwell.
- Tønnessen, Finn Egil. 2002. 'ICT and communication (problems)'. In S. Strömqvist (ed.), *The diversity of languages and language learning*, 91-101. Lund: Lund University.

Per Henning Uppstad <Per_Henning.Uppstad@ling.lu.se>

Listeners' sensitivity to consonant variation within words

Joost van de Weijer

1 Introduction

Part of our native language competence is the implicit knowledge of phonological word structure. Speakers of English know that *flink* is a possible word of English, but that *lfink* is not, because phonotactic constraints do not permit the combination *lf* as a word onset.

Experimental work with infants shows that this knowledge develops at a very early age. Jusczyk et al. 1993, for instance, demonstrated that infants at the age of nine months have knowledge of the sounds that occur in their native language. In this experiment, a group of American infants and a group of Dutch infants were tested using the preferential looking paradigm. The infants in both groups listened to Dutch and English words. The words were matched in terms of word length and stress pattern, but the Dutch words contained speech sounds that are not part of the English sound system, whereas the English words contained speech sounds that are not part of the Dutch sound system. At nine months of age, the American infants had a listening preference for the English words, and the Dutch infants had a listening preference for the Dutch words. At six months of age, there was no difference between the two groups, suggesting that sensitivity to the sound system of the native language develops between six and nine months of age.

In another study, it was shown that infants in the same period develop knowledge of phonotactic patterns in their native language (Jusczyk, Luce & Charles-Luce 1994). The infants in this study listened to nonsense words that contained combinations of speech sounds that were either highly probable or highly unlikely in their native language. The results of this study were similar to those of Jusczyk et al. 1993. Nine-month-old infants preferred to listen to stimuli with high-probable sound combinations but six-month-old infants had no preference yet.