Word-formation types
Definition, identification, classification

Andreas Nolda
http://andreas.nolda.org
Szegedi Tudományegyetem
Német Nyelvészeti Tanszék

Magyar Tudományos Akadémia
Nyelvtudományi Intézet
23 January 2014

How not to define word-formation types (I)

Bauer (2006: 719) attempts to 'define compounds' as follows:

> A compound is usually defined (somewhat paradoxically) as a word that is made up of two other words. This basic definition requires a certain amount of modification, some of it for all languages, some of it for specific languages. For example, there may be more than two 'words' involved in the formation of a compound, though there must be at least two. […] the forms in which the individual subwords appear may be differently defined in different languages: a citation form in one, a stem in another, a specific compounding form in yet a third, a word form in a fourth. […] Perhaps the rider should be added that the construction created by the two or more lexemes should not be a normal noncompound phrasal structure of the language […].

How not to define word-formation types (II)

In my view, a 'definition of compounds' along Bauer's lines is misconceived in several respects:

– What should be sought for, are nominal definitions of terms like "compounding" and "compound", explicating the general intuitions linguists have about compounding, compounds etc. (Cf. Hempel 1952 for the general distinction between *nominal definitions*, *real definitions*, and *explications* in philosophy of science.)

– The non-phrasehood – i.e. wordhood – of word-formation products should not be a definition criterion for terms like "compound", but fall out as a theorem of a properly constructed word-formation theory.

– Definitions of theoretical terms like "compounding" and "compound" should be universal, not specific for a linguistic system. Otherwise such terms would not be applicable, in principle, to all languages in the same way.

– Whether a unit in a given linguistic system is to be analysed as, say, a compound (and thereby as a word) is a question of identifying the entities in that linguistic system which fall under the concept denoted by the term "compound" as defined in the presupposed word-formation theory, which, typically, requires additional, system-specific, criteria. (For the linguistic distinction between *definition* and *identification*, cf. Budde 2000.)

How to define word-formation types (I)

In the following, I shall show how terms like "compounding" and "compound" can be defined in a properly constructed word-formation theory, namely the *Pattern-and-Restriction Theory* of word formation (*PR*; Nolda 2012).

PR is an axiomatically formalised word-formation theory in the Item-and-Process tradition, with the following basic ideas:

– The *word-formation component* (*WF component*) of a linguistic system provides the *word-formation patterns* (*WF patterns*) in the system and their *base restrictions*.

– WF patterns are multidimensional constructs from formal, categorial, and semantic operations, the *word-formation means* (*WF means*).

– *Word-formation processes* (*WF processes*) are modelled as functions from WF patterns to word-formation operations on multidimensional elements in the restriction of the pattern.

– WF processes serve to establish *word-formation relations* (*WF relations*) on lexicalised ('existing') as well as non-lexicalised ('possible') lexical units in the system.

The design of PR has been influenced in particular by Aronoff's (1976) word-formation rules and morphological restrictions, Beard's (1995) Lexeme-Morpheme Base Morphology and earlier versions of Lieb's (2013) Process Model of Word Formation. The formalisation of PR presupposes the general theoretical framework of Integrational Linguistics (Lieb 1983, 1999; Budde 2000; Nolda 2007).

How to define word-formation types (II)

In PR, definitions for terms like "compounding" and "compound" are introduced in the following logical order:

1. Terms like "compounding" are defined as certain types of WF processes, by reference to (classes of) WF patterns in the WF component of the linguistic system.

2. Terms like "compound" are defined as the products of the corresponding types of WF processes in the linguistic system.

In this way, necessary and sufficient criteria are provided for distinguishing between word-formation types, given a linguistic system and its word-formation component.

The domain of exemplification will be compounding in Modern German, with an outlook to other types of word formation.[1]

1 The theoretical framework in a nutshell

Lexical and lexicological units (I)

Lexical units – *lexical words*, *lexical stems*, etc. – are conceived as pairings of a paradigm of forms and a lexical meaning.

Notational conventions:

lexical meanings: "'settlement'"

lexical words: "$Ort^{\mathrm{W}}_{\text{'settlement'}}$"

lexical stems: "$Ort^{\mathrm{St}}_{\text{'settlement'}}$"

---

[1] The core of PR and some sample WF patterns in Modern German have been implemented in a computer program called "PPR" ('processing system for word-formation patterns and their restrictions'), which can be downloaded from http://andreas.nolda.org/index.php/software.

Lexical and lexicological units (II)

Lexical units can be categorised by sets of *lexical categories*, such as part-of-speech categories or nominal gender categories in languages like German:

$Ort^{\text{W}}_{\text{`settlement'}}$: {Noun, Masc}

$Ort^{\text{St}}_{\text{`settlement'}}$: {NounSt, MascSt}

Lexical and lexicological units (III)

Formally and semantically related lexical units are members of the same *lexicological unit* (*lexicological words*, *lexicological stems*):

(1)   a. $Ort^{\text{W}}_{\text{`place'}}$

      b. $Ort^{\text{W}}_{\text{`settlement'}}$

(2)   a. $Ort^{\text{St}}_{\text{`place'}}$

      b. $Ort^{\text{St}}_{\text{`settlement'}}$

Forms and their categories (I)

Word forms are modelled as sequences (possibly, unit sequences) of syntactic atoms, and *stems forms* as sequences of morphological atoms.

Syntactic and *morphological atoms* are phonological constructs linking a segmental unit with suprasegmental representations, determining, in particular, lexical accents.
    Notational conventions:

primary lexical accents: "ˈ"

secondary lexical accents: "ˌ"

atoms: "ˈ*ort*"

Forms and their categories (II)

Word forms of $Ort^{\text{W}}_{\text{`settlement'}}$:

(3)   a.   ˈ*ort*

      b.   ˈ*ort(e)s*

      c.   ˈ*orts*

      d.   ˈ*orten*

   Stem forms of $Ort^{\text{St}}_{\text{`settlement'}}$:

(4)   a.   ˈ*ort*

      b.   ˈ*ort e*

      c.   ˈ*ort s*

      d.   ˈ*ört*

Forms and their categories (III)

A *paradigm* categorises the forms by sets of *paradigmatic categories*.

The paradigm of $Ort^{\mathrm{W}}_{\text{`settlement'}}$ provides the following word-form categorisations:

$\quad$ 'ort: $\{\text{Sing-Nlf}, \text{Nom/Acc/Dat/Gen-Nlf}\}$

$\quad$ 'ort(e)s: $\{\text{Sing-Nlf}, \text{Gen-Nlf}\}$

$\quad$ 'orte: $\{\text{Plur-Nlf}, \text{Nom/Acc/Dat/Gen-Nlf}\}$

$\quad$ 'orten: $\{\text{Plur-Nlf}, \text{Dat-Nlf}\}$

Nom/Acc/Dat/Gen-Nlf is a default category, which can be overwritten by more specific categories such as Gen-Nlf or Dat-Nlf.

The paradigm of $Ort^{\mathrm{St}}_{\text{`settlement'}}$ provides the following stem-form categorisations:

$\quad$ 'ort: $\{\text{NlStf}\}$

$\quad$ 'ort e: $\{\text{Plur-NlStf}\}$

$\quad$ 'ort s: $\{\text{Comp-NlStf}\}$

$\quad$ 'ort: $\{\text{Der-NlStf}\}$

$\quad$ 'ört: $\{\text{Der-NlStf}\}$

NlStf is again a default category.

WF relations and WF processes (I)

(5) exemplifies a WF relation in Modern German systems **S**:

(5) $\quad$ $Kur\ ort^{\mathrm{St}}_{\text{`health-cure settlement'}}$ is formed from $Kur^{\mathrm{St}}_{\text{`health cure'}}$ and $Ort^{\mathrm{St}}_{\text{`settlement'}}$, through two-place compounding in **S** ($\mathrm{comp}^2_{\mathbf{S}}$).

$\quad$ (5) is a *direct WF relation* in **S**, because certain formal, categorial, and semantic properties of $Kur\ ort^{\mathrm{St}}_{\text{`health-cure settlement'}}$ can be directly determined on the basis of properties of $Kur^{\mathrm{St}}_{\text{`health cure'}}$ and $Ort^{\mathrm{St}}_{\text{`settlement'}}$ through $\mathrm{comp}^2_{\mathbf{S}}$:

1. The stem forms 'kur ˌort, 'kur ˌort e, etc. of $Kur\ ort^{\mathrm{St}}_{\text{`health-cure settlement'}}$ result from concatenating the stem form 'kur of $Kur^{\mathrm{St}}_{\text{`health-cure'}}$ with deaccented variants of the stem forms 'ort, 'ort e, etc. of $Ort^{\mathrm{St}}_{\text{`settlement'}}$.

2. The paradigmatic categorisations of the stem forms of $Kur\ ort^{\mathrm{St}}_{\text{`health-cure settlement'}}$ are identical to those of the stem forms of $Ort^{\mathrm{St}}_{\text{`settlement'}}$.

WF relations and WF processes (II)

3. The lexical categorisation of $Kur\ ort^{\mathrm{St}}_{\text{`health-cure settlement'}}$ is identical to that of $Ort^{\mathrm{St}}_{\text{`settlement'}}$.

4. The lexical meaning 'health-cure settlement' of $Kur\ ort^{\mathrm{St}}_{\text{`health-cure settlement'}}$ is a partially idiomatised variant of the concept 'settlement in relation to health cure(s)', which can be constructed by combining the lexical meanings of $Kur^{\mathrm{St}}_{\text{`health cure'}}$ and $Ort^{\mathrm{St}}_{\text{`settlement'}}$ in an appropriate way.

$\quad$ Due to the direct WF relation (5) between the lexical stems $Kur\ ort^{\mathrm{St}}_{\text{`health-cure settlement'}}$, $Kur^{\mathrm{St}}_{\text{`health cure'}}$, and $Ort^{\mathrm{St}}_{\text{`settlement'}}$, there holds also an *indirect WF relation* between the corresponding lexical words:

(6) $\quad$ $Kurort^{\mathrm{W}}_{\text{`health-cure settlement'}}$ is formed from $Kur^{\mathrm{W}}_{\text{`health cure'}}$ and $Ort^{\mathrm{W}}_{\text{`settlement'}}$ through $\mathrm{comp}^2_{\mathbf{S}}$.

WF patterns and base restrictions (I)

Those properties are determined through the WF process by means of a *WF pattern*, by applying it to elements of its *base restriction*.

WF patterns are modelled as quadruples from operations of the following sort:

1. a *formal WF means*, operating on forms

2. a *paradigmatic WF means*, operating on sets of paradigmatic categories

3. a *lexical WF means*, operating on sets of lexical categories

4. a *semantic WF means*, operating on sets of lexical meanings

All WF means in a pattern are operations of the same arity, which is the arity of the pattern as a whole.

WF patterns and base restrictions (II)

The WF pattern relevant for relation (5) is a quadruple of two-place operations:

(7)   ⟨concat-after-deacc-of-arg2,
          ident-with-arg2,
          ident-with-arg2,
          rel-of-arg2-and-arg1⟩

WF patterns and base restrictions (III)

Taken together, (5) specifies a two-place operation on multidimensional quadruples with a formal, a paradigmatic, a lexical, and a semantic component, to be called *word-formation instances* (*WF instances*) of lexical units:

(8)   a.   ⟨ˈkur,
                {NlStf},
                {NounSt, FemSt},
                'health cure'⟩

        b.   ⟨ˈort,
                {NlStf},
                {NounSt, MascSt},
                'settlement'⟩

        c.   ⟨ˈkur ˌort,
                {NlStf},
                {NounSt, MascSt},
                'health-cure settlement'⟩

*WF patterns and base restrictions (IV)*

The base restriction of a WF pattern provides those WF instances to which the pattern may apply:

– The base restriction of a one-place WF pattern is a set of WF instances.

– The base restriction of a two-place WF pattern is a set of pairs of WF instances.

…

WF patterns and base restrictions (V)

The value of the operation specified by pattern (7) for the pair $\langle(8\,a),(8\,b)\rangle$ in its base restriction is (8 c):

$$
(9) \quad \left(\mathrm{comp}_{\mathsf{S}}^{2} \left(\begin{matrix} \langle \text{concat-after-deacc-of-arg2,} \\ \text{ident-with-arg2,} \\ \text{ident-with-arg2,} \\ \text{rel-of-arg2-and-arg1} \rangle \end{matrix} \right) \right) \left(\begin{matrix} \langle \textprime kur, \\ \{\mathrm{NlStf}\}, \\ \{\mathrm{NounSt, FemSt}\}, \\ \text{`health cure'}), \\ \langle \textprime ort, \\ \{\mathrm{NlStf}\}, \\ \{\mathrm{NounSt, MascSt}\}, \\ \text{`settlement'}) \end{matrix} \right)
$$

$$
= \quad \begin{matrix} \langle \textprime kur\, {}_{\backprime}ort, \\ \{\mathrm{NlStf}\}, \\ \{\mathrm{NounSt, MascSt}\}, \\ \text{`health-cure settlement'}) \end{matrix}
$$

Thereby, $\mathrm{comp}_{\mathsf{S}}^{2}$ determines formal, categorial, and semantic properties of $Kur\ ort^{\mathrm{St}}_{\text{`health-cure settlement'}}$, on the basis of properties of $Kur^{\mathrm{St}}_{\text{`health cure'}}$ and $Ort^{\mathrm{St}}_{\text{`settlement'}}$.

2 General word-formation types

General approach

Terms like "two-place compounding in linguistic systems S" ($\mathrm{comp}_{\mathsf{S}}^{2}$) are defined as certain types of WF processes, with reference to properties of the WF patterns in their domain.

Those properties will be determined with reference to classes in a classification system on the WF patterns of a linguistic system, established by criteria of different sort.

Logical criteria

(10)

$$
\begin{array}{c}
\text{WF pattern in } \mathbf{S} \\
| \\
\textbf{general arity}
\end{array}
$$

one-place WF pattern in **S**   multi-place WF pattern in **S**

multi-place arity

two-place WF pattern in **S**   three-place WF pattern in **S**   …
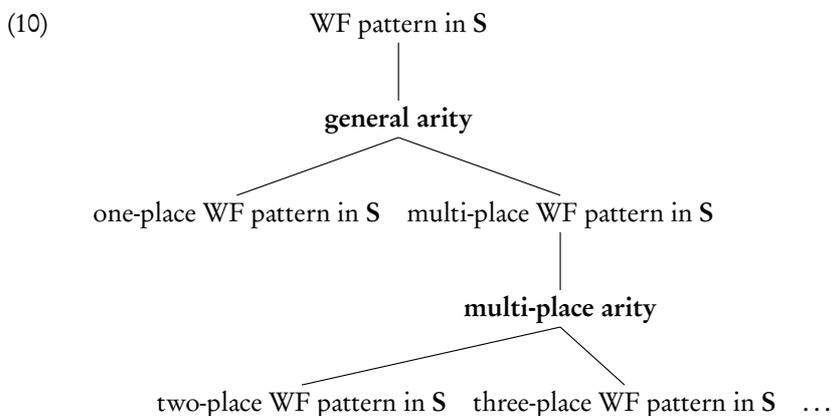
"General arity" is the name of a first classification on the WF patterns in Modern German systems **S**. Its classification basis is the class 'WF pattern in **S**'. The classes in this classification are 'one-place WF pattern in **S**' and 'multi-place WF pattern in **S**', which in turn functions as the basis of the classification named "multi-place arity". In general, classes in a classification may overlap; their union is the classification basis (Juilland and Lieb 1968).

6

Compounding and derivation

The general arity of WF patterns in the domain of WF processes serves to distinguish between:

general derivation: word formation with one-place WF patterns

general compounding: word formation with multi-place WF patterns

(11) a. *Definition*
      general derivation in S (gen-der$_S$) = the WF process in S whose domain consists of all one-place WF patterns in S

  b. *Definition scheme*
      Let $n \geq 2$.
      n-place general compounding in S (gen-comp$_S^n$) = the WF process in S whose domain consists of all n-place WF patterns in S

By the definition scheme (11 b), comp$_S^2$, comp$_S^3$, etc. are defined. Compounding processes with arities > 2 are used for forming copulative compounds like *rot gelb grün*$_{\text{'red-yellow-green'}}^{\text{St}}$.

Formal criteria

(12)   WF pattern in **S**

            |

            general tactical change

            /   \

        fusional   non-fusional
      WF pattern in **S**   WF pattern in **S**
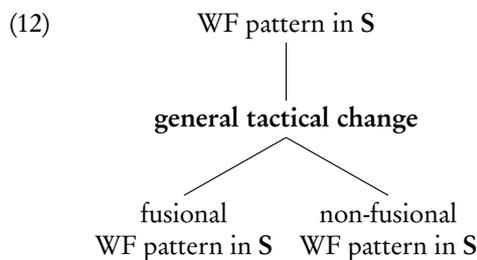
Fusional patterns are those with a formal means which combine several morpho-syntactic atoms forms into one:

(13) overlap('kur, 'urlaub) = 'kurlaub

Non-fusional patterns have formal means which do not combine atoms into one:

(14) concat-after-deacc-of-arg2('kur, 'urlaub) = 'kur $_{\text{,}}\text{urlaub}$

Compounding and blending

In the view taken here, blending and (proper) compounding are both subtypes of general compounding (cf. also Donalies 2005). They can be distinguished by formal criteria only:

(proper) compounding: word formation with multi-place non-fusional WF patterns

blending: word formation with multi-place fusional WF patterns

(15) a. *Definition*
      Let $n \geq 2$.
      n-place (proper) compounding in S (comp$_S^n$) = the WF process in S whose domain consists of all n-place non-fusional WF patterns in S

  b. *Definition*
      Let $n \geq 2$.
      n-place blending in S (blend$_S^n$) = the WF process in S whose domain consists of all n-place fusional WF patterns in S

Compounds

Terms for word-formation products like "(proper) compound" can be defined along the following lines:

(16)   *Definition*
        l is a *(proper) compound* in S if, and only if, there is an n, $l_1, \ldots, l_n$ such that l is formed from $l_1, \ldots, l_n$ through comp_S^n.

Semantic criteria (I)

Endocentric and exocentric compounds, and endocentric and exocentric word-formation products in general, are usually semantically distinguished (cf., for instance, Bloomfield 1933: 235):

(17)   a.   *Metall zylinder*$_{\text{'metal cylinder'}}^{\text{St}}$

        b.   *Sechs zylinder*$_{\text{'object with six cylinders'}}^{\text{St}}$

The semantic criterion to be used here is *base implication*:

(18)

WF pattern in **S**

base implication

base-implying   non-base-implying
WF pattern in **S**   WF pattern in **S**
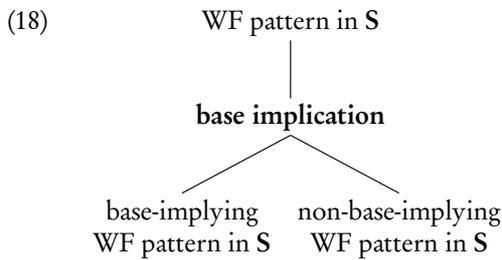
Semantic criteria (II)

(19)   a.   *Definition*
              $\langle \phi^n, \pi^n, \lambda^n, \sigma^n \rangle$ is an *n-place base-implying WF pattern* if, and only if,
              1. $\langle \phi^n, \pi^n, \lambda^n, \sigma^n \rangle$ is an n-place WF pattern in S and
              2. there is an i with $1 \leq i \leq n$ such that each value of σ^n implies the i-th argument of σ^n.

        b.   *Definition*
              $\langle \phi^n, \pi^n, \lambda^n, \sigma^n \rangle$ is an *n-place non-base-implying WF pattern* if, and only if,
              1. $\langle \phi^n, \pi^n, \lambda^n, \sigma^n \rangle$ is an n-place WF pattern in S and
              2. there is no i with $1 \leq i \leq n$ such that each value of σ^n implies the i-th argument of σ^n.

Endocentric and exocentric word formation

  endocentric word formation: word formation with base-implying WF patterns

  exocentric word formation: word formation with non-base-implying WF patterns

(20)   a.   *Definition*
              n-place endocentric word formation in S (endo-wf_S^n) = the WF process in S whose domain consists of all *n*-place base-implying WF patterns in S

        b.   *Definition*
              n-place exocentric word formation in S (exo-wf_S^n) = the WF process in S whose domain consists of all *n*-place exocentric WF patterns in S

8

Semantic criteria (III)

The next criteria serves to determine patterns for coordinative compounds:

(21) a. *Öst er reich ungarn*$_{\text{'Autria-Hungary'}}^{\text{St}}$

   b. *Kur urlaub*$_{\text{'health-cure-holiday'}}^{\text{St}}$

   c. *Kurlaub*$_{\text{'health-cure-holiday'}}^{\text{St}}$

(22)   multi-place WF pattern in **S**

                     |

            commutativity

         ╱   ╲

   commutative   non-commutative
   WF pattern in **S**   WF pattern in **S**
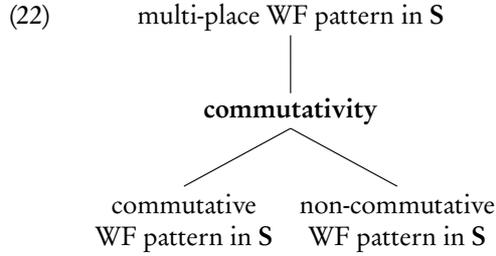
Semantic criteria (IV)

(23) *Definition*
Let $n \geq 2$.
$\langle \phi^n, \pi^n, \lambda^n, \sigma^n \rangle$ is an *n-place commutative WF pattern* if, and only if,
   1. $\langle \phi^n, \pi^n, \lambda^n, \sigma^n \rangle$ is an *n*-place WF pattern in S and
   2. for all $\langle c_1, \ldots, c_n \rangle$ in the domain of σ^n, $\sigma^n(c_1, \ldots, c_n) = \sigma^n(c_n, c_1, \ldots) = \ldots$

(24) a. $\text{sum}(\text{'Austria'}, \text{'Hungary'}) = \text{sum}(\text{'Hungary'}, \text{'Austria'})$

   b. $\text{conj}(\text{'health cure'}, \text{'holiday'}) = \text{conj}(\text{'holiday'}, \text{'health cure'})$

Coordinative and subordinative compounding

   coordinative general compounding: general compounding with commutative WF patterns

   subordinative general compounding: general compounding with non-commutative WF patterns

(25) a. *Definition*
   n-place coordinative general compounding in S (coord-gen-comp$_S^n$) = the WF process in S whose domain consists of all multi-place commutative WF patterns in S

   b. *Definition*
   n-place subordinative general compounding in S (sub-gen-comp$_S^n$) = the WF process in S whose domain consists of all multi-place non-commutative WF patterns in S

As a rule, sub-comp$_S^n$ is two-place.

3 Conclusion

Definition of terms for word-formation types

Terms like "general compounding", "(proper) compounding", "subordinative compounding" etc. were defined as certain types of WF processes, by reference to logical, formal, and semantic properties of WF patterns in the WF component of the linguistic system.

Those properties were determined with reference to classes in a classification system on the WF patterns in a linguistic system.

Definition of terms for types of word-formation products

Terms like "(proper) compound" are definable as the products of the corresponding types of WF processes in the linguistic system.

Identification of word-formation products

In principle, the word-formation products in a linguistic system are (partially or totally) identified by identifying the WF patterns and their base restrictions in the system.

This, however, is not a theoretical tasks, but an empirical one.

References

Aronoff, Mark (1976). *Word Formation in Generative Grammar*. Linguistic Inquiry Monographs 1. Cambridge, Mass.: MIT Press.

Bauer, Laurie (2006). Compound. In *The Encyclopedia of Language and Linguistics*, ed. by Keith Brown *et al.*, 2nd ed., Amsterdam: Elsevier, vol. 2, 719–726.

Beard, Robert (1995). *Lexeme Morpheme Base Morphology: A General Theory of Inflection and Word Formation*. Albany: State University of New York Press.

Bloomfield, Leonard (1933). *Language*. New York: Holt, Rinehart and Winston.

Budde, Monika (2000). Wortarten: Definition und Identifikation. Ph.D. thesis, Freie Universität Berlin.

Donalies, Elke (2005). *Die Wortbildung des Deutschen: Ein Überblick*. 2nd ed. Studien zur deutschen Sprache 27. Tübingen: Narr.

Hempel, Carl G. (1952). *Fundamentals of Concept Formation in Empirical Science*. Chicago: University of Chicago Press.

Juilland, Alphonse G. and Hans-Heinrich Lieb (1968). *„Klasse" und Klassifikation in der Sprachwissenschaft*. Janua Linguarum: Series Minor 74. The Hague: Mouton.

Lieb, Hans-Heinrich (1983). *Integrational Linguistics*. Current Issues in Linguistic Theory 17. Amsterdam: Benjamins. Vol. 1: *General Outline*.

Lieb, Hans-Heinrich (1999). Morph, Wort, Silbe: Umrisse einer Integrativen Phonologie des Deutschen. In *Variation und Stabilität in der Wortstruktur: Untersuchungen zu Entwicklung, Erwerb und Varietäten des Deutschen und anderer Sprachen*, ed. by Matthias Butt and Nanna Fuhrhop, Germanistische Linguistik: Monographien 141/142, Hildesheim: Olms, 334–407.

Lieb, Hans-Heinrich (2013). Towards a general theory of word formation: The Process Model. Freie Universität Berlin. http://edocs.fu-berlin.de/docs/receive/FUDOCS_document_000000018561 [5 Aug. 2013].

Nolda, Andreas (2007). *Die Thema-Integration: Syntax und Semantik der ‚gespaltenen Topikalisierung' im Deutschen*. Studien zur deutschen Grammatik 72. Tübingen: Stauffenburg.

Nolda, Andreas (2012). Konversion im Deutschen – Muster und Beschränkungen: Mit einem Grundriss einer allgemeinen Theorie der Wortbildung. Habilitation thesis, Humboldt-Universität zu Berlin.