1. Introduction

This article is concerned with two types of truncatory processes: truncated personal names as they are used in many languages to form vocatives and hypocoristics, and truncated non-names. For ease of reference, I will use the term 'truncated names' to refer to the former and 'clippings' to refer to the latter. Examples of the two types of process are given in (1) and (2).

(1) a. German

---

1 I would like to thank the participants of the Workshop Irregularity in Morphology, held at the Annual Meeting of the Societas Linguistica Europaea, Bremen, Aug 30 - Sept 2, 2006, as well as audiences at the first International Conference on the Linguistics of Contemporary English, Edinburgh, June 23-26, 2005, the meeting of the DFG network Core Mechanisms of Exponence, Leipzig, Jan 11-12, 2008, and the Sprachwissenschaftliches und sprachdidaktisches Kolloquium at the University of Siegen, Dec 12, 2007, where earlier versions of this paper were presented. Special thanks are due to Birgit Alber, for many fruitful discussions of the cross-linguistic structural properties of truncation, which helped greatly to sharpen my view on the morphological aspects of the process discussed in this paper, and to Ingo Plag, for providing enlightening and constructive feedback during various stages of the development of this paper. Finally, I would like to thank the editors of this volume, especially Elke Ronneberger-Sibold, for her patience and for stimulating discussions of the pertinent theoretical issues. Needless to say, the usual disclaimers apply.
In terms of the topic of this volume, truncatory processes constitute an interesting case because their status as instances of regular word formation is debated in the literature. In much of the descriptive literature on word-formation systems in languages as well as in recent studies from a functional perspective, it is assumed that truncatory processes are somehow different from regular word-formation. In much work in morphological theory, then, not much attention is devoted to truncation. One notable exception is recent work set within the framework of Natural Morphology, where a strong emphasis is put on the distinction between

---

2 This view is, for example, reflected in textbooks on morphological theory such as Haspelmath (2002: 25), Booij (2005: 21).
what is termed grammatical operations of inflection and word formation on the one hand and extragrammatical operations on the other hand (Dressler 2000), (Dressler 2005), (Ronneberger-Sibold in prep). Together with other processes like sound symbolism, blending, and word manufacture, truncation as a whole is viewed as an extragrammatical or creative operation (Ronneberger-Sibold 2010, Ronneberger-Sibold in prep).

Many of the properties that have led to the assessment of the status of truncatory processes as outside morphological grammar can be subsumed under the central notion of predictability. Specifically, two aspects of truncation are often assumed to be unpredictable in the literature. The first concerns variability of the shape of the output (cf. esp. Ronneberger-Sibold 2010 for a recent version). The assumption underlying this claim is that 'speakers are free to leave the beaten track and modify or mix the old techniques, thus coming up with new solutions [...] tailored to the problem at hand' (p. 204). The second aspect concerns the relation between the output of truncation and its base. Given that most truncated words lack an internal morphemic structure, and that they contain only a fragment of their bases, they are, by definition, less transparent than outputs of other, concatenative morphological processes, and are sometimes even considered to be non-transparent. Given their low degree of transparency, then, it is assumed that the relation between bases and truncated forms is unpredictable in the sense that bases are not recoverable from truncated words. An interesting proposal in this respect is that non-recoverability of the base form does in fact provide a functional motivation for truncation (Ronneberger-Sibold 2001).

In the present paper I will present several case studies, many of which are taken from my empirical work on English (Lappe 2003, Lappe 2005, Lappe 2007), that test the unpredictability claim against empirical evidence from two types of truncatory processes that are found frequently in the world's languages: truncated personal names and clippings.

With respect to the structural properties of the investigated truncatory patterns, I will argue that unpredictability is not a helpful concept to describe what is going on in truncation, and that a more adequate analysis of the investigated patterns crucially involves making a distinction between systematic and highly constrained variability of structural characteristics on the one hand and isolated idiosyncracies on the other. If this distinction is made, the idea that truncation is fundamentally different from other grammatical processes, turns out to be highly problematic.

With respect to base recoverability, I will discuss relevant structural properties of truncations in the light of psycholinguistic evidence about factors that facilitate or impede word

---

3 Apart from predictability, further criteria that have been used to assess the status of truncation are the criterion of intentionality and semantic and / or pragmatic criteria. Cf. Alber & Arndt-Lappe (to appear) for a summary.
recognition and lexical access. The basic idea is that, unlike in other morphological processes, in truncation base recovery is essentially word recognition from a fragment. I will argue that in name truncation and clipping we find strong evidence that, despite the loss of transparency, the truncatory patterns under investigation strive to preserve recoverability. It will thus be shown that the claim that truncation is motivated by the aim to obscure recoverability cannot be extended to all patterns of truncation. In more general terms, the study of truncation sheds an interesting light on the question how transparency is in fact related to recoverability.

In general, the theoretical implication of the evidence discussed in this paper is that there do exist in the world predictable processes of truncation. Hence, it is impossible to draw a sharp line between grammatical and extragrammatical morphology, as is done in Natural Morphology. Note, however, that this does not preclude the possibility that some patterns are more predictable than others. Thus, for example, one would assume that the trade names discussed by Ronneberger-Sibold in various places (esp. Ronneberger-Sibold 2010, Ronneberger-Sibold in prep) have a different status than, for example, truncated personal names which are regularly used to form hypocoristics in languages such as English (Lappe 2003, Lappe 2005, Lappe 2007) or Italian (Alber 2007, Alber 2010). Again, this observation then constitutes evidence that no sharp line can be drawn between truncation as grammatical word-formation and truncation as extragrammatical operation.

2. Structural Predictability

2.1 The literature

A general problem in assessing the predictability of output structures of truncatory processes is that, in much of the pertinent literature, the degree of predictability is not in the focus of investigation. A further problem is that major work on truncation comes from very different theoretical frameworks, which happen to make radically different claims about the predictability of truncatory patterns. In what follows, the situation will be illustrated mostly for English truncation, one of the most well studied languages in this respect.

For English, there is a large descriptive tradition that assumes that output structure of truncation is unpredictable. This view is found, e.g., in classic treatments of English word-formation such as (Jespersen 1965repr) or (Marchand 1960), and in some of the major textbooks in this area (e.g. Bauer 1983). The same view is also found in work set in Natural Morphology (Dressler & Merlini Barbaresi 1994, Dressler 2000, Dressler 2005, Ronneberger-Sibold 2010, Ronneberger-Sibold in prep), focussing on languages other than English. The
key evidence for the unpredictability claim comes from base forms for which several different truncated forms are attested. Oft-cited examples are given in (3).

(3) Beth, Bess, ... < Elisabeth (Dressler & Merlino Barbaresi 1994)
    Bill < William (Jespersen 1965 repr)

The view that truncation is structurally unpredictable has, however, come under attack from work in formal phonological theory, most notably from work set in the framework of Prosodic Morphology (McCarthy & Prince 1986 [1996], McCarthy & Prince 1993 et seq.). In this research program, truncatory patterns have often been cited as evidence for the claim that prosodic categories (esp. the syllable and the metrical foot) play an important role in determining the structure of outputs of morphology. The clearest case documenting the influence of prosodic categories on truncatory outputs is the case in which the truncated form corresponds exactly to such a prosodic category in a one-to-one fashion. This is exactly what was found in a number of studies inspired by Prosodic Morphology: Many languages have truncatory patterns in which the output of truncation regularly corresponds to a metrical foot of that language. This invariant output shape has either been modelled as a prosodically prespecified template (cf. esp. McCarthy & Prince 1986 [1996]: 45ff., Weeda 1992 for a cross-linguistic overview) or, within the framework of optimality-theoretic Generalised Template Theory (McCarthy & Prince 1999, with roots in McCarthy & Prince 1994), as the result of constraint interaction. For English, a templatic analysis was proposed in (Weeda 1992); a constraint-based account can be found in (Lappe 2003), (Lappe 2005), (Lappe 2007).

One problem with many studies in Prosodic Morphology is that, given the interest of the research program, their focus is necessarily limited. Thus, several studies look only at prosodically determined size restrictions (e.g. Bat-El 2002 for Hebrew, Downing 2006 for various languages). Much more rarely discussed is the question which part of the base is retained in the truncation (e.g. Piñeros 1998 for Spanish, Féry 1997 for German i-formations). The same is true for effects such as segmental changes and substitution phenomena, which are often cited as evidence for the alleged unpredictability of truncation (cf. e.g. [w] and [b] in the pair William and Bill in (3) above), but which are often not investigated in detail.

In terms of the formal issues covered, however, studies set in the Prosodic Morphology paradigm have uncovered a high degree of formal regularity in truncatory patterns of several different languages (cf. e.g., Alber 2007, Alber 2010 for Italian and German, Bat-El 2002 for Hebrew, Féry 1997, Wiese 2001 for German, Lappe 2003, Lappe 2005, Lappe 2007 for
Based on a large, cross-linguistic survey of documented patterns, (Alber & Arndt-Lappe 2007-2009) and (Alber & Arndt-Lappe to appear) conclude that the unpredictability claim is untenable. They propose a formal model to account for the typology of truncatory structures set within the research paradigm of Prosodic Morphology.

In sum, we see that the literature on truncation is characterised by conflicting claims w.r.t. the structural predictability of truncated forms. Before we can, however, evaluate these claims against empirical data, we need to clarify which levels of structural analysis need to be addressed. I will distinguish between three levels of analysis, for which I will use the following labels (cf. also Lappe 2007):

- word structure
- anchoring
- segmental stability

'Word structure' here refers to the overall structure of truncated forms, measured in terms of the number of syllables and the stress pattern. The term 'anchoring' addresses the question of which part of the base word is retained in the truncation. Both word structure and anchoring have traditionally been claimed to be notoriously unpredictable in English truncation. Thus, for example, Bauer writes in his 1983 textbook on English word-formation:

It does not seem to be predictable how many syllables will be retained in the clipped form (except that there are fewer than in the base lexeme), whether the final syllable will be open or closed, whether the stressed syllable from the base lexeme will be included or not. While a clipping which retains the initial part of the word is easily the most common type, there are also others. (Bauer 1983: 233; my emphasis)

A similar view is found in Kreidler's 1979 article on clipping:

Nobody can predict what terms are likely to be so shortened nor how the shortened forms will 'catch on' and spread. Nor can the precise shape of the shortening be predicted, where the cut will be made, how much retained, how much discarded. However, investigation shows that some possibilities are greater than others. There are various probabilities when big words are creatively shortened. (Kreidler 2000: 28f.; emphasis added)

Finally, I will use the term 'segmental stability' to address the question in how far the sounds of the base word that are retained in the truncated word are retained faithfully. Like unpredictability of word structure and anchoring, segmental instability has been traditionally attributed to truncated words in English. The most influential reference here seems to be (Jespersen 1965repr: 533ff.), who devotes quite a large part of his discussion of shortenings to the issue of the 'numerous irregularities' (p. 540) that he sees in truncated forms, and most of which he attributes to 'the speech habits of small children' (p. 540), who in his view are the major creators of truncated forms.
In what follows I will present case studies that investigate aspects of formal predictability for truncatory patterns pertaining to the three levels of analysis mentioned. In section 2.2 we will first look at word structure and anchoring in three sets of data: English name truncation and clippings, Italian name truncation, and German name truncation. In section 2.3 I will present a case study on segmental stability in two patterns of English truncation.

### 2.2 Word structure and anchoring

#### 2.2.1 English name truncation and clipping

The data as well as the analysis are taken from (Lappe 2007). Table 1 provides an overview of the origin of the data to be cited in this paper.

<table>
<thead>
<tr>
<th>name truncation</th>
<th>Number of different base-derivative pairs:</th>
</tr>
</thead>
<tbody>
<tr>
<td>• a private U.S. website set up as a resource for genealogical research in Franklin County, Tennessee (Phillips 2002)</td>
<td>948</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>word clipping</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>• all forms labelled 'colloquial abbreviation', extracted from the electronic Second Edition of the (1994) (OED)</td>
<td>702</td>
</tr>
<tr>
<td>• a dictionary of American English slang (Spears 1989)</td>
<td></td>
</tr>
<tr>
<td>• a dictionary of English slang from different varieties (Beale 1989)</td>
<td></td>
</tr>
</tbody>
</table>

Table 1. Data from (Lappe 2007), to be cited in this paper; adapted from: (Lappe 2007: 59f.)

All data have been collected for a purpose independent of the study of truncatory morphology. For name truncation, the overwhelming majority of names in the corpus are names of people who have lived in a well-defined geographical region (Franklin County, TN). For clipping, the corpus is more heterogeneous, mostly comprising colloquial slang words.

An analysis of word structures of truncated names and clippings in my corpus shows that different patterns are attested for the truncation of names and non-names (i.e. clippings), respectively. An overview of attested word structures is given in Tables 2 and 3. Strong, i.e. stressed syllables are represented as 'S', weak syllables are represented as 'w'. In addition, cases in which final segments are analysed as suffixes in (Lappe 2007) are marked as such.

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>%</th>
<th>examples:</th>
</tr>
</thead>
<tbody>
<tr>
<td>S</td>
<td>386</td>
<td>40.72%</td>
<td>Rube (&lt; Ruben)</td>
</tr>
<tr>
<td>S w, y-suffixed</td>
<td>380</td>
<td>40.08%</td>
<td>Minnie (&lt; Minerva)</td>
</tr>
<tr>
<td>S w, [ə]-suffixed</td>
<td>135</td>
<td>14.24%</td>
<td>Rena (&lt; Irene)</td>
</tr>
<tr>
<td>other</td>
<td>47</td>
<td>4.96%</td>
<td>Arilla (&lt; Cinderella)</td>
</tr>
</tbody>
</table>
Table 2. Word structure in truncated names in English name truncation

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>%</th>
<th>examples:</th>
</tr>
</thead>
<tbody>
<tr>
<td>S w</td>
<td>421</td>
<td>60.0%</td>
<td>mish (&lt; missionary)</td>
</tr>
<tr>
<td>S w, y-suffixed</td>
<td>94</td>
<td>13.4%</td>
<td>assy (&lt; asphalt)</td>
</tr>
<tr>
<td>S w, o-suffixed</td>
<td>99</td>
<td>14.1%</td>
<td>dero (&lt; derelict)</td>
</tr>
<tr>
<td>w S</td>
<td>47</td>
<td>6.7%</td>
<td>exec (&lt; executive)</td>
</tr>
<tr>
<td>other</td>
<td>41</td>
<td>5.8%</td>
<td>influ (&lt; influenza)</td>
</tr>
<tr>
<td><strong>Σ</strong></td>
<td>702</td>
<td>100.0%</td>
<td></td>
</tr>
</tbody>
</table>

Table 3. Word structure in clippings in English clipping

For name truncation, three clear word structure patterns emerge which together account for about 95% of the corpus data (40.72% + 40.08% + 14.24%). Of these three patterns, the only pattern that does not involve a suffix is the monosyllabic pattern, which comprises a single, bimoraic syllable. Both disyllabic patterns can be analysed as suffixed.4

For clipping, four systematic word structure patterns are attested in the data, which together account for some 94% of all data. Two of these patterns are suffixed, involving the suffixes –y and –o, which have both been described as regular diminutive and/or depreciative suffixes in the literature (cf. esp. Wierzbicka 1984 for a semantic analysis). Like for name truncation, the most frequent unsuffixed clipping type is the bimoraic monosyllable. Unlike name truncations, however, unsuffixed clippings may also be disyllabic, consisting of an iambic word structure (e.g. exéc < exécutive, celéb < celébrity). Clippings of this type are attested for cases where the base word begins with an unstressed syllable that is followed by a stressed syllable.

For a survey of anchoring patterns we will now look at each of the word structure patterns in turn. For convenience, these are summarised in (4) and (5). Note that [ə]-suffixed forms and unsuffixed iambic clippings will not be discussed any further.

(4) unsuffixed patterns

a. names: Pete < Peter

---

4 Note that the suffix –y has different spelling variants (esp. –ie, cf. Schneider (2003)), which all represent the same phonological form, [i]. For ease of reference, I will use the letter –y to refer to this suffix. The main basis for the analysis of –y and -[ə] as suffixes is that both -[i] and -[ə] occur in forms in which [i] and [ə] are not etymological, i.e. do not occur in pertinent positions in the base words. The assumption that –[i] in English truncations is a suffix is a conventional assumption made in the pertinent literature. Cf., e.g., Schneider (2003), Dressler & Merlini Barbaresi (1994), Poynton (1989), Wierzbicka (1984) for discussion.
b. clippings, monosyllabic: mike < microphone
    fab < fabulous
    oc < octopus

(5) suffixed patterns
  a. y-suffixed names:  Pety < Peter
                        Genie < Eugene
                        Trishy < Patricia
  b. y-suffixed clippings: ciggy < cigarette
                           hammy < hamster
                           veggie < vegetable
  c. o-suffixed clippings: evo < evening
                           delo < delegate
                           aggro < aggravation

Tables 4 and 5 summarise anchoring patterns for monosyllabic truncated names and clippings. To make the distribution of the two emerging patterns clearer, the tables show anchoring patterns only for those base words where main stress is non-initial.

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>%</th>
<th>examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>initial syllable</td>
<td>63</td>
<td>53.0%</td>
<td><em>Hez</em> (&lt; Hezekiah)</td>
</tr>
<tr>
<td>main-stressed syllable</td>
<td>50</td>
<td>42.0%</td>
<td><em>Kye</em> (&lt; Hezekiah)</td>
</tr>
<tr>
<td>other</td>
<td>6</td>
<td>5.0%</td>
<td><em>Beth</em> (&lt; Elizabeth)</td>
</tr>
<tr>
<td><strong>Σ</strong></td>
<td>119</td>
<td>100.0%</td>
<td></td>
</tr>
</tbody>
</table>

Table 4. anchoring in monosyllabic names for bases where main stress is non-initial

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>%</th>
<th>examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>initial syllable</td>
<td>123</td>
<td>90.4%</td>
<td><em>ack</em> (&lt; acknowledge)</td>
</tr>
<tr>
<td>main-stressed syllable</td>
<td>10</td>
<td>7.4%</td>
<td><em>sheen</em> (&lt; machine)</td>
</tr>
<tr>
<td>other</td>
<td>3</td>
<td>2.2%</td>
<td><em>droid</em> (&lt; android)</td>
</tr>
<tr>
<td><strong>Σ</strong></td>
<td>137</td>
<td>100.0%</td>
<td></td>
</tr>
</tbody>
</table>

Table 5. anchoring in monosyllabic clippings for bases where main stress is non-initial
The tables show two things about anchoring in monosyllabic truncations: First of all, truncated names and clippings show different anchoring properties, a fact that provides further support for the idea that these two types of truncation should be kept apart in the analysis. Secondly, within the two types, we find that anchoring is highly regular. Thus, about 95% of all truncated names in the corpus anchor to either the first or the main-stressed syllable of their bases. Among clippings, about 90% anchor to the initial syllable.\(^5\)

Tables 6 and 7 show that the anchoring patterns found among monosyllabic truncations are mirrored by those found among \(y\)-suffixed forms.

<table>
<thead>
<tr>
<th></th>
<th>(N)</th>
<th>%</th>
<th>examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>initial syllable</td>
<td>195</td>
<td>55.6%</td>
<td>\textit{Sibbie} (&lt; \textit{Sibilla})</td>
</tr>
<tr>
<td>main-stressed syllable</td>
<td>146</td>
<td>41.6%</td>
<td>\textit{Zaddy} (&lt; \textit{Arzada})</td>
</tr>
<tr>
<td>other</td>
<td>10</td>
<td>2.8%</td>
<td>\textit{Betty} (&lt; \textit{Elisabeth})</td>
</tr>
<tr>
<td>(\Sigma)</td>
<td>351</td>
<td>100.0%</td>
<td></td>
</tr>
</tbody>
</table>

Table 6. Anchoring in \(y\)-suffixed truncated names for bases where main stress is non-initial

<table>
<thead>
<tr>
<th></th>
<th>(N)</th>
<th>%</th>
<th>examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>initial syllable</td>
<td>40</td>
<td>93.0%</td>
<td>\textit{chrissie} (&lt; \textit{chrysanthemum})</td>
</tr>
<tr>
<td>other</td>
<td>3</td>
<td>7%</td>
<td>\textit{toxie} (&lt; \textit{intoxicated})</td>
</tr>
<tr>
<td>(\Sigma)</td>
<td>43</td>
<td>100.0%</td>
<td></td>
</tr>
</tbody>
</table>

Table 7. Anchoring in \(y\)-suffixed clippings for bases where main stress is non-initial

We see the familiar distribution, which is, however, even clearer for suffixed than for unsuffixed cases: Some 97% of all \(y\)-suffixed names anchor to either the initial or the main-stressed syllable of their bases, whereas among clippings only initial anchoring seems to be an option. This option is chosen by some 93% of the data in the corpus.\(^6\)

Finally, Table 8 provides an overview of anchoring in \(o\)-suffixed clippings, for which there is no parallel pattern in name truncation.

<table>
<thead>
<tr>
<th></th>
<th>(N)</th>
<th>%</th>
<th>examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>initial syllable</td>
<td>66</td>
<td>98.5%</td>
<td>\textit{aggro} (&lt; \textit{aggravation})</td>
</tr>
<tr>
<td>other</td>
<td>1</td>
<td>1.5%</td>
<td>\textit{Frisco} (&lt; \textit{San Francisco})</td>
</tr>
<tr>
<td>(\Sigma)</td>
<td>43</td>
<td>100.0%</td>
<td></td>
</tr>
</tbody>
</table>

Table 8. Anchoring in \(o\)-suffixed clippings for bases where main stress is non-initial

---

\(^5\) Similar numbers emerge if we look at truncated words which are derived from bases where main stress is initial: Here 86.3% of all names anchor to that syllable, compared to 91.8% of all clippings.

\(^6\) Again, the distributions among truncations based on words with initial stress are fully in line with these findings: Here 91.8% of all names and 98% of all clippings anchor to the word-initial syllable.
With only one exception, all pertinent forms in the corpus anchor to the initial syllable of their bases.  

In sum, the data investigated in Lappe's (2007) study suggest that output structure and anchoring in name truncation and clipping is predictable in English. This claim is based on the fact that in the corpora, we find that the overwhelming majority of truncated names and clippings (a) falls into a small number of distinct patterns which (b) can be clearly defined in terms of their structural properties, some of which are (c) specific to their pattern. As an example of the latter generalisation, consider how truncated names differ from clippings:

Monosyllabic and \( y \)-suffixed name truncations and clippings show different anchoring patterns; iambic truncations are possible for clippings, but ungrammatical for names.

An important point is that this interpretation of the empirical facts is based on the underlying assumption that the small number of truncations that cannot be classified into the patterns described are idiosyncratic, exceptional cases. It is precisely this assumption where the approach proposed here differs from other approaches, which would interpret the existence of cases deviating from the majority pattern as evidence that the majority patterns reflect tendencies, not regular rules of word-formation. The supporting evidence for the view adopted in this paper is twofold: First of all there is the sheer strength of the tendencies found in the corpora, and the very small number of exceptional cases. Secondly, there is reason to assume that for each idiosyncratic form in the corpus there is also an alternative form that is formed according to the regular pattern. As an example, consider the 47 base-derivative pairs where the word structure of the derivative has been classified as 'other' in Table 2, which comprise truncated names for 43 different base names. For 28 of these 43 base names (i.e. for about 65%), the same database also contains at least one fully regular truncated form. 'Fully regular' here means that the form follows the regular pattern in terms of both anchoring and word structure, and does not differ from its base in any unsystematic way in terms of its segmental makeup. Examples of regular and irregular forms for the same bases are given in (6).

\[
\begin{array}{lll}
\text{base} & \text{regular form(s)} & \text{irregular form(s)} \\
Elizabeth & Liz, Lizzy & Bess, Beth, Bet, Betty, Ibbly, Libby, ... \\
Abraham & Abe & Bram \\
Sylvester & Syl, Vest & Sli, Si, Vester \\
Achilles & Kill & Killis \\
\end{array}
\]

\(^7\) For bases with initial main stress, the percentage is 100%.
Examples of base names for which only irregular truncated forms are attested include Alexander (Andi, Sandy, Al, Alec, Alex), Bazaleel (Basil), Bedelia (Delia), Griselda (Grissel), Leonard (Leo). For these cases I assume that the fact that no regular form is included in the corpus is accidental.

2.2.2 Italian name truncation

In what follows I will describe word structure and anchoring patterns as found in a sample of Italian name truncations that were collected by Birgit Alber and myself from undergraduate students participating in a linguistics course at the universities of Verona and Trento (date: Nov 18 & 19, 2002). The students, which are all speakers of Northern varieties of Italian, were asked to supply (in written form) truncated forms of names of people they know. The survey yielded 244 different base-derivative pairs (orthographic variants were ignored). A detailed structural analysis can be found in (Alber 2007), (Alber 2010), which is also the major reference on which I base my categorisation of word structure and anchoring patterns below.8 Note that it is not the goal of this section to provide a full account of the structural characteristics of Italian name truncation, but merely to show which word structure patterns are attested in general. Specifically, the analysis of word structure provided below does not do justice to the fact that Italian also has productive reduplicated truncation (cf. Alber 2007, Alber 2010 for a detailed account). If they have the same word structure, reduplicated and non-reduplicated forms are not distinguished.

Table 9 provides an overview of word structures in the data, using the format already familiar from the discussion of English.

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>%</th>
<th>examples:</th>
</tr>
</thead>
<tbody>
<tr>
<td>S w</td>
<td>110</td>
<td>45.1%</td>
<td>Anto (&lt; Antonella)</td>
</tr>
<tr>
<td>S w, i-final</td>
<td>94</td>
<td>38.5%</td>
<td>Andri (&lt; Andrea)</td>
</tr>
<tr>
<td>S</td>
<td>34</td>
<td>13.9%</td>
<td>Giò (&lt; Giovanni)</td>
</tr>
<tr>
<td>w S</td>
<td>5</td>
<td>2.0%</td>
<td>Milé (&lt; Milena)</td>
</tr>
</tbody>
</table>

8 Another detailed structural analysis of Italian truncatory patterns can be found in Thornton (1996), Thornton (2004). Given that I will limit the discussion to our specific sample of names, which all originate from speakers of Northern varieties of Italian, I will not discuss Thornton's account here.
Whereas the clarity with which distinct patterns emerge from the data is comparable to what we saw for English, we also see that despite some similarities, Italian name truncations also markedly differ from English patterns. With one single exception, all data fall into four distinct patterns: There are obviously two highly frequent disyllabic patterns, one of which takes two syllables from the base word and ends in an open syllable, and one of which involves a fixed segment, -i, which may be analysed as a suffix. The latter is apparently similar to the relevant English pattern (but cf. the differences in the behaviour of intervocalic consonant clusters: compare Italian Andri ˂ Andrea to English Andy ˂ Andrew). The former, however, is not found in English, where the predominant simplex pattern is monosyllabic. In Italian there is also a monosyllabic pattern, which differs, however, from the similar English pattern in that this syllable is usually codaless. Finally, in Italian there seems to be a pattern which shows iambic stress, a pattern that in English we find only among clippings, not among truncated names. The status of the iambic pattern in Table 9 is, however, unclear. Four of the five bases of the forms attested in the data are stressed on the second syllable (Maria, Milēna, Morēno, Sofía). It is thus unclear whether these forms are part of a disyllabic iambic pattern, or whether they are exponents of a pattern reported by in (Alber 2007), (Alber 2010) for Southern varieties of Italian, where the truncated form preserves the stretch from the initial to the main-stressed syllable of the base, regardless of the number of syllables involved (cf., e.g., Antoné ˂ Antonélla).

Tables 10 - 12 look at anchoring in the three major Italian patterns: the unsuffixed disyllabic pattern ('Sw-truncations'), the i-final disyllabic pattern ('i-final Sw-truncations'), and the monosyllabic pattern. As in the discussion of the English data, only bases are considered in which main stress is non-initial.

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>%</th>
<th>examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>initial syllable</td>
<td>51</td>
<td>48.57%</td>
<td>Marghe (˂ Margherita)</td>
</tr>
<tr>
<td>main-stressed syllable</td>
<td>49</td>
<td>46.67%</td>
<td>Betta (˂ Elisabetta)</td>
</tr>
</tbody>
</table>

9 All truncated forms ending in –i were coded as 'i-final' for the frequency analysis provided in Table 9. This means that among the data coded as 'i-final', some –i s may also be interpreted to be etymological (e.g. Vanni ˂ Giovanni). Note that the decision about how to code forms with etymological –i does not bear on the main point to be made here, namely that most Italian truncated names have a disyllabic, trochaic structure.
Table 10. Anchoring in Sw-truncations for bases where main stress is non-initial

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>%</th>
<th>examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>initial syllable</td>
<td>69</td>
<td>87.34%</td>
<td>Albi (&lt; Alberto)</td>
</tr>
<tr>
<td>main-stressed syllable, noninitial</td>
<td>9</td>
<td>11.39%</td>
<td>Resy (&lt; Teresa)</td>
</tr>
<tr>
<td>unclear</td>
<td>1</td>
<td>1.27%</td>
<td>Tessy (&lt; Stefania)</td>
</tr>
<tr>
<td>Σ</td>
<td>79</td>
<td>100.0%</td>
<td></td>
</tr>
</tbody>
</table>

Table 11. Anchoring in i-final Sw-truncations for bases where main stress is non-initial

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>%</th>
<th>examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>initial syllable</td>
<td>17</td>
<td>94.44%</td>
<td>Ste (&lt; Stefania)</td>
</tr>
<tr>
<td>main-stressed syllable, noninitial</td>
<td>1</td>
<td>5.56%</td>
<td>Cè (&lt; Francesca)</td>
</tr>
<tr>
<td>Σ</td>
<td>18</td>
<td>100.0%</td>
<td></td>
</tr>
</tbody>
</table>

Table 12. Anchoring in monosyllabic truncations for bases where main stress is non-initial

Like in the English data, anchoring is surprisingly uniform, with an even smaller percentage of irregular or unclear forms than in the English dataset. In principle Italian displays the same anchoring types as English, with almost all forms anchoring to either the initial or the main-stressed syllable of their bases. However, the distribution of anchoring types among truncatory patterns is markedly different. Simplex (i.e. unsuffixed) disyllabic names may anchor to the initial or the main-stressed syllable of their bases. Simplex monosyllabic name truncations, however, and i-suffixed disyllables seem to almost only anchor to the initial syllable of their bases, and not to the main-stressed syllable. They are thus more similar to English clippings, which show the same pattern of preference, than to English truncated names, for which both anchoring patterns are equally important.

2.2.3 German name truncation

There is quite a large body of literature that discusses the structural properties of major patterns of German truncation (cf., e.g., Ronneberger-Sibold 1992, Ronneberger-Sibold 1995, Féry 1997, Steinhauer 2000, Steinhauer 2007, Wiese 2001, Köpcke 2002, Alber 2007); especially the pattern of i-suffixed truncation in both name truncation and word clipping has
attracted much interest. As in the discussion of Italian name truncation in section 2.2.2, it can
thus not be the goal of the present discussion to provide a full account of the structural
properties of German name truncation. Like the discussion of the Italian data, the present
discussion complements existing research in that it provides an empirical survey of word
structure and anchoring in a sample of data. The sample was collected in the summer term of
2007 from university students at the University of Siegen, who attended an introductory
course in linguistic pragmatics, and who were asked to supply nicknames and full names of
people they know personally. The survey yielded 568 different base-derivative pairs.10
Table 13 provides an overview of the distribution of word structure patterns. As in the other
datasets, I assume that final segments which occur as non-etymological segments in some of
the data can be analysed as suffixes. In these cases, all forms ending in these segments are
classified as 'x-final', regardless of the etymological status of the final segment, and treated as
a separate pattern in the table.

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>%</th>
<th>examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>S w</td>
<td>144</td>
<td>25.4%</td>
<td>Karo (&lt; Karolin)</td>
</tr>
<tr>
<td>S</td>
<td>78</td>
<td>13.7%</td>
<td>Jo (&lt; Johann)</td>
</tr>
<tr>
<td>S w, i-final</td>
<td>285</td>
<td>50.2%</td>
<td>Wolfi (&lt; Wolfgang)</td>
</tr>
<tr>
<td>S w, e-final</td>
<td>36</td>
<td>6.3%</td>
<td>Wolle (&lt; Wolfgang)</td>
</tr>
<tr>
<td>S w, e(r)l-final</td>
<td>13</td>
<td>2.3%</td>
<td>Naddel (&lt; Nadine)</td>
</tr>
<tr>
<td>chen-final diminutives</td>
<td>8</td>
<td>1.5%</td>
<td>Inchen (&lt; Ina)</td>
</tr>
<tr>
<td>other</td>
<td>4</td>
<td>0.7%</td>
<td>Elisa (&lt; Elisabeth)</td>
</tr>
<tr>
<td>Σ</td>
<td>568</td>
<td>100.0%</td>
<td></td>
</tr>
</tbody>
</table>

Table 13. Word structure in the German data

We see that i-suffixed forms, i.e. the structure that has received most attention in the
literature, make up about half of the data (50.2%) provided by the German students. Apart
from i-suffixed forms, however, German has at least two other, unsuffixed patterns of name
truncation: These are unsuffixed disyllables where the final syllable is an open syllable and
unsuffixed monosyllables, which together account for some 39% of the data. What is also
remarkable is that the i-suffixed pattern is not the only suffixed pattern. In the data we also
find a set of 36 e-final forms (where –e is phonologically realised as [ə]), which fulfills the

10 Note that some data were were excluded from further analysis because there was either no or only an
idiosyncratic structural relation between the base and the derivative (e.g. Jensemann < Jens, Pünktchen < Anton),
or because the nickname was a (pseudo-) anglicised version of the base name (e.g. Kässi < Kathrin).
criterion for suffixed patterns because in some of the relevant forms –e is clearly non-etymological (compare, e.g., Wolle < Wolfgang, Ede < Edmund, Edgar, or Domme < Dominik). Finally, we find in the data a small number of forms which involve suffixes which are have traditionally not been analysed as truncatory suffixes (-chen, -e(r)l), and a very small number of truncated forms which do not correspond to any of the major word-structure patterns. As in the other languages discussed, the latter make up only a very small proportion of the data (0.4%).

Tables 14 – 16 provide an overview of anchoring patterns found in the three major patterns attested in the data: unsuffixed monosyllabic and disyllabic forms (the latter will be referred to as 'Sw-truncations'), and i-suffixed disyllabic forms (i-suffixed Sw-truncations). As in the previous discussion, the relevant dataset will be restricted to those forms where main stress in the base form is non-initial. Note also that e-final disyllables will not be discussed any further – closer examination of the structural properties of this pattern will be a task for future research.

### Table 14. Anchoring in unsuffixed Sw-truncations for bases where main stress is non-initial

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>%</th>
<th>examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>initial syllable</td>
<td>42</td>
<td>38.5%</td>
<td>Manu (&lt; Manuéla)</td>
</tr>
<tr>
<td>main-stressed syllable, noninitial</td>
<td>66</td>
<td>60.6%</td>
<td>Ela (&lt; Manuéla)</td>
</tr>
<tr>
<td>unclear</td>
<td>1</td>
<td>0.1%</td>
<td>Nina (&lt; Kristína)</td>
</tr>
<tr>
<td><strong>Σ</strong></td>
<td>109</td>
<td>100.0%</td>
<td></td>
</tr>
</tbody>
</table>

### Table 15. Anchoring in i-suffixed Sw-truncations for bases where main stress is non-initial

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>%</th>
<th>examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>initial syllable</td>
<td>96</td>
<td>72.7%</td>
<td>Brunni (&lt; Brunhilde)</td>
</tr>
<tr>
<td>main-stressed syllable, non-initial</td>
<td>35</td>
<td>26.5%</td>
<td>Tini (&lt; Bettína)</td>
</tr>
<tr>
<td>unclear</td>
<td>1</td>
<td>1.0%</td>
<td>Betty (&lt; Elisabeth)</td>
</tr>
<tr>
<td><strong>Σ</strong></td>
<td>132</td>
<td>100.0%</td>
<td></td>
</tr>
</tbody>
</table>

### Table 16. Anchoring in e-final disyllabic forms

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>%</th>
<th>examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>initial syllable</td>
<td>14</td>
<td>77.8%</td>
<td>Lu (&lt; Luísa)</td>
</tr>
<tr>
<td>main-stressed syllable, non-initial</td>
<td>4</td>
<td>22.2%</td>
<td>Nel (&lt; Cornélia)</td>
</tr>
<tr>
<td>unclear</td>
<td>0</td>
<td>0.0%</td>
<td></td>
</tr>
</tbody>
</table>

16
As in the English and the Italian data, anchoring to another but the initial or the main-stressed syllable is highly exceptional. Note also that there are anchoring differences between languages. Thus, both German and Italian have an unsuffixed disyllabic pattern of name truncation. Whereas, however, Italian truncations with this structure may anchor to either the initial or the main-stressed syllable, German truncations with this structure seem to show a preference for main-stress anchoring.

### 2.3 Segmental stability

This section presents a small case study on English whose purpose is to assess the claim frequently made in the literature that, on the segmental level, truncation often involves irregular, i.e. unpredictable phenomena (cf., e.g., Jespersen 1965repr: 541 for a formulation of the problem). I will discuss the consonant stability of English monosyllabic patterns: truncated names and clippings. The analysis, which will again be based on (Lappe 2007), will show two things: First of all, the vast majority of truncated forms does not involve unpredictable segmental phenomena. Secondly, looking at attested truncations, a distinction should be made between on the one hand segmental phenomena which are indeed unpredictable and idiosyncratic, and, on the other hand, segmental phenomena which are systematic and predictable. I assume that in a corpus, idiosyncratic segmental phenomena manifest themselves by occurring in a few, isolated forms, but, crucially, do not occur in all pertinent contexts. By contrast, regular segmental phenomena may occur in all pertinent contexts. For regular segmental phenomena we will find that they are often optional. We will also see, however, that their optionality depends on the truncatory pattern.

Table 17 looks at the stability of consonantal segments in English monosyllabic truncated names in my corpus (Lappe 2007). 'Stability' here refers to the relation between the base name and the truncated form.

<table>
<thead>
<tr>
<th>N</th>
<th>%</th>
<th>examples:</th>
</tr>
</thead>
</table>

1. Among base-derivative pairs where main stress in the base is word-initial, the distribution looks as follows: 91.6% of all monosyllabic truncations, 85.3% of all unsuffixed disyllabic truncations, and 88.2% of all i-suffixed disyllabic truncations anchor to the initial/main-stressed syllable of their bases.

2. The selection of consonantal changes and of monosyllabic patterns here is arbitrary. The same point could be made for vowels and for other patterns. Cf. Lappe (2007) for the pertinent details.
We see that in the vast majority of truncated names (88.6%) consonants in the truncated name are fully identical to the relevant consonants in their bases. Among the segmental changes observed, one group of 17 forms stands out as involving a systematic change: the dental fricative [θ] in the base name corresponds to [t] in the truncated name. Closer investigation reveals that the change is systematic and optional. I assume that the change is systematic because, for all 20 pertinent base forms in the database, a truncated name is attested in the same database that has [t] where the base has [θ]. The optionality of the segmental change is evidenced by the fact that for each of the pertinent bases, an alternant form can be found that faithfully retains [θ]. Within the database, this is true for three bases (Bartholomew: Bart, Barth, Elisabeth: Beth, Bette, Thaddeus: Thad, Tad). A wider search in the internet reveals that this is true also for the other bases in the corpus. Examples are provided in (7).

(7) optionality of segmental changes involving the dental fricative

<table>
<thead>
<tr>
<th>base name</th>
<th>consonant change</th>
<th>no consonant change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Martha</td>
<td>Mart</td>
<td>Marth</td>
</tr>
<tr>
<td>Cynthia</td>
<td>Cynt</td>
<td>Cynth</td>
</tr>
<tr>
<td>Bertha</td>
<td>Bert</td>
<td>Berth</td>
</tr>
<tr>
<td>Nathaniel</td>
<td>Nat</td>
<td>Nath</td>
</tr>
</tbody>
</table>

Finally, we see in Table 17 that other consonant changes are attested for 27 base-derivative pairs in my 2007 corpus. The forms are discussed in detail in (Lappe 2007: 99f.). I consider them to be idiosyncratic because, in all cases, the change seems to be restricted to isolated base-derivative pairs (or small sets of base-derivative pairs, cf. below). Crucially, for all forms involving the change alternant forms can be found that do not display the change. What is interesting about many of these changes is that they often occur in only one very specific truncated name, which may, however, be used as a truncated name for several different bases. This latter observation may leave the impression that the change is indeed systematic in some sense. However, there is good evidence that the nature of this systematicity is very different from the systematic cases that we discussed further above. A
good example is the truncated name Bill, which is traditionally used for the base William, but which is also attested for bases like Willis and Wilbert. The case that [w] in the base corresponds to [b] in the truncated name occurs only if the output is the form Bill, and, crucially, not in other contexts. For example, Winfield, Wendy, and Webster cannot become *Bin, *Bend, or *Beb, but only (and regularly) Win, Wen, and Web.\textsuperscript{13} I conclude that correspondence between [w] in the base and [b] in the truncation is different from the type of correspondence that was labelled systematic above and that was discussed for the case of the [θ]-[t] alternation. In the latter case we are dealing with a morpho-phonological regularity in the traditional sense (that can be captured in terms of rules or constraints) that applies to a general phonological context (that can be conceptualised, e.g., as a phonological feature or group of features). In the former case the alternation does not apply to a general phonological context of this type, but to one single, very specific one (i.e. a particular truncated name).\textsuperscript{14} Having looked at segmental stability of consonants in name truncation, we now turn to the same issue in clipping. Table 18 provides the relevant distributions.

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>%</th>
<th>examples:</th>
</tr>
</thead>
<tbody>
<tr>
<td>no change</td>
<td>410</td>
<td>97.6%</td>
<td>mensh (&lt; mention)</td>
</tr>
<tr>
<td>[θ] &gt; [t]</td>
<td>0</td>
<td>0.0%</td>
<td>maths (&lt; mathematics)</td>
</tr>
<tr>
<td>other</td>
<td>10</td>
<td>2.4%</td>
<td>bike (&lt; bicycle)</td>
</tr>
<tr>
<td>Σ</td>
<td>421</td>
<td>100.0%</td>
<td></td>
</tr>
</tbody>
</table>

Table 18. Consonant stability in clipping

We see that in clipping, the number of segmental changes in consonants is even smaller than in name truncation. 97.6% of the data contain the same consonants as the relevant portion of their bases. There are only 10 base-derivative pairs which exhibit a change, and none of the changes appears to be systematic. What is interesting is that the dental fricative, which regularly undergoes a change in name truncation, seems to be fully stable in clipping. The pertinent forms from the corpus are given in (8).

(8) stability of [θ] in clipping

\textsuperscript{13} Note that the reason cannot be homonymy avoidance, because, like bin and bend, also win, when, and web are attested words of English.

\textsuperscript{14} The effect that truncated forms sometimes become associated with base names to which they are not related etymologically has already been observed by Karl Sundén in his 1904 dissertation on name truncation in English (or, in his terminology, 'elliptical personal names', cf. Sundén (1904: 140ff.). He termed this phenomenon 'pseudoellipsis', and argued convincingly on the basis of historical records that pseudoellipsis is the source of many apparent irregularities among lexicalised truncated names.
maths  (< mathematics)
cath  (< catheter)
meth  (< methedrine)
synth (< synthesiser)
thesp (< thespian)

Like the anchoring data discussed in section 2.2., the data on consonant changes in English name truncation and clipping therefore show that the exact details of the structural properties of truncatory outputs depend on the relevant pattern. We have also seen that these properties can be described in terms of systematic phonological restrictions on output structure and on the relation between bases and derivatives. Thus, the observed differences between patterns as well as the systematicity of the changes observed constitute evidence that truncation is not much different from other morphological processes: Like other morphological processes, truncation comes with a cophonology\textsuperscript{15} (Orgun 1996, Inkelas & Orgun 1998, and subsequent work), i.e. a set of phonological restrictions that holds for outputs of a particular morphological process, but not necessarily across the language in general.

3. Base Recoverability

3.1 Introduction

The assumption that transparency is a characteristic of productive morphological processes is commonplace in much of the morphological literature (cf. e.g. Ronneberger-Sibold 2001, Braun & Plag 2003). Nevertheless, we find that the term is used with slightly different senses in the literature. In the psycholinguistic literature on morphological processing the term transparency is often employed to refer to the degree to which morphemes in a morphologically complex word are formally and semantically related to the base morphemes from which they derive (cf., e.g., Libben et al. 2003 for discussion). In the theoretical morphological literature we find that in some approaches also the aspect of compositionality, i.e. predictability of meaning resulting from the combination of morphemes, is important. For example, Ronneberger-Sibold defines transparency as 'the possibility of inferring a meaning from the parts of such a word or phrase and the way they are combined.' (Ronneberger-Sibold 2001: 98, my emphasis).

\textsuperscript{15} Nothing in the argument proposed here hinges on the particular framework adopted. The cophonology approach is only one of several approaches that have been proposed to capture the observed facts.
According to this latter view, then, truncatory processes are not transparent by definition because here formal compositionality does not correspond to semantic compositionality. As an example, consider the English truncated name *Ed*, derived from the base *Edward*. We may argue that *Ed* has a diminutive meaning component and, thus, differs semantically from *Edward* (cf. esp. Schneider 2003, Alber & Arndt-Lappe to appear for discussion). The meaning of *Ed* as a complex meaning is, however, not reflected in the form *Ed*, which is clearly not compositional.

However, in terms of a definition of transparency as a measure of recoverability of bases within the truncation, the issue of how transparent outputs of truncation are becomes less trivial. Crucially, then, base recoverability in truncation is a process that must be very similar to word recognition from word fragments. It is this type of transparency that I will discuss in this section, relating, where possible, relevant structural properties of truncation to pertinent findings from psycholinguistic research. To avoid ambiguity, I will henceforth use the term 'base recoverability' to refer to the phenomenon in truncation. The purpose of this section is to show that rules governing the formation of truncated words in many patterns of truncation seem to be geared towards facilitating recoverability of the base.

The crucial question to be discussed here is: In how far does the reduction in form that we find in outputs of truncation obscure base recoverability? Since there is to my knowledge no empirical work that addresses this question directly, the discussion will have to be confined here to indirect evidence. Sections 3.2, 3.3, and 3.4 will be devoted to three different sources of such evidence. In section 3.2 we will look at anchoring in truncation and relate the attested productive patterns to findings that have emerged from the psycholinguistic literature about properties of words that play a role in lexical access and word recognition. In section 3.3 we will then discuss the problem that, because of the reduction of phonological form, truncation may lead to homonymous truncated forms for different base words. A large number of homonymous truncations can be seen as a factor obstructing base recoverability because, given homonymous truncations, speakers cannot know which is the right base form. For ease of reference, I will henceforth refer to this problem as 'the homonymy problem'. For English name truncation and clipping, we will take stock of the distribution of homonymous forms in the English truncation corpora from (Lappe 2005), (Lappe 2007). Finally, in section 3.4 we

---

16 This is particularly true in types of truncation which do not involve a suffix.
17 Note that I will confine the argument here to showing that there is indeed a strong relation between bases and truncated forms that has the potential of to make base recovery happen. The question of whether or not the base is in fact recovered whenever a derived word occurs is not addressed here.
will discuss some preliminary evidence about how recoverable bases are when truncations are
used in discourse context. The pertinent data will again come from English.

3.2 Anchoring and word recognition

The empirical basis of the argument developed in this section is a cross-linguistic survey of
structural generalisations and anchoring patterns in truncation that was done by the present
author in collaboration with Birgit Alber (Alber & Arndt-Lappe 2007-2009, Alber & Arndt-
Lappe to appear).

An important problem that we face when investigating cross-linguistic regularities in
anchoring patterns is that empirical studies systematically investigate anchoring, and which
distinguish between different truncatory patterns are rare. Thus, many studies within Prosodic
Morphology have tended to be selective in terms of the phenomena studied, and the focus has
often been on word structure rather than on anchoring. Among studies set in other
frameworks, we often find that the aim is classification of different types rather than
providing a quantitative account. And, again, among the few existing quantitative studies,
many do not investigate anchoring regularities (e.g. Ronneberger-Sibold 1992 on German and
French).

Still, even in studies that do not explicitly investigate anchoring regularities, we often find
that authors comment on observed generalisations. Examples of pertinent languages are given
in (9) – (11).

(9) French (Scullen 1997)

a. abrèv < abréviation
docu < documentaire
capi < capitaine

b. pitaine < capitaine
péca < ipekakwana (a Tropical plant)

(10) Hebrew templatic hypocoristics (Bat-El 2005)

a. méni < menáxem
?ási < ?ásaf
tiki < tikva

b. náxi, xémi < menáxem
sáfi < ?ásaf

(11) Kopfwörter and Endwörter in Swedish (Nübling 2001)
In the studies cited in (9) – (11) the underlying assumption is that anchoring is variable and unsystematic. Still, we find that authors comment on observed anchoring regularities, which raise questions about how unsystematic (in the sense discussed in section 2) anchoring really is in the patterns cited. Thus, for French clippings Scullen (1997) basically assumes that 'establishing a single site for the mapping of elements to a template [...] appears to be futile' (p. 97). Still, she admits that left anchoring (in her terminology: 'mapping from the left edge', p. 97) is 'the standard case' (p. 97). A similar comment can be found in Bat-El's (2005) study of Hebrew hypocoristics. Focussing on regularities in output structure of truncated names, she notes that 'THs [templatic hypocoristics, S.A.-L.] come in various forms when their correspondence to their base is considered: left-anchored, misanchored [sic!], and reduplicated, again, either left-anchored or misanchored.' (p. 126, my emphasis). Finally, in her comparative study of German and Swedish truncations Nübling (2001) notes that, in principle, different anchoring patterns are attested, those where the initial part of the base lexeme is retained ('Kopfwörter'), and those where the final part is retained ('Endwörter'). For the latter type, however, she notes:


The examples cited indicate that, even in studies which do not focus explicitly on anchoring or which do not presuppose that anchoring is systematic, it has frequently been noted that anchoring does not appear to be arbitrary, but that there are at least very strong tendencies. In the case studies on English name truncation and clipping, and German and Italian name truncation that were discussed in section 2.2 we saw the same phenomenon. Whereas truncatory patterns differ in terms of which anchoring types they allow, anchoring is surprisingly uniform. We find that all patterns allow anchoring to material that is initial in the base. Additionally, some patterns allow anchoring to material that is main-stressed in the base. Together, initial anchoring and main-stress anchoring account for more than 90% of all data collected in these pilot studies.

(Alber & Arndt-Lappe 2007-2009), (Alber & Arndt-Lappe to appear) systematically investigate which anchoring patterns (and word structure patterns) are attested cross-
linguistically. The survey is based on published work on truncation set in different frameworks as well as on our own empirical work on Italian, German, and English. It comprises 91 truncation patterns from 27 different languages. The findings show that what we saw in the pilots in section 2.2, can well be extended to other languages: The predominant anchoring pattern is left-edge anchoring (attested in 50.5% of all truncation patterns in the database), followed by main-stress anchoring (attested in 16.5% of all pertinent truncation patterns). In addition, there are patterns which preserve both initial and main-stressed material from their bases (7.7% of all patterns in the database).

In sum, there is converging evidence that the overwhelming majority of truncatory patterns preserves initial and to main-stressed material of their bases. Furthermore, there is evidence that anchoring to base-initial material is more widespread than anchoring to main-stressed material. These two facts are interesting because they can be directly related to well-known findings in the psycholinguistic literature about the role of initial material and stress-related information in word recognition (cf. Beckman 1998, Alber 2001 for similar arguments within phonological theory).

The central role of word-initial material in word recognition has been demonstrated in experimental research across different languages and using an array of different tasks. Thus, for example, in experiments in which participants are put into a tip-of-the-tongue state, it turns out that initial parts of words are among those18 that speakers are more likely to remember than other parts (cf. Brown & McNeill 1966 for a classic experiment). There is, furthermore, a type of psycholinguistic experiments that is very relevant to our present discussion, because the task almost simulates the task that hearers of truncated words face, of decoding bases. This type of experiments investigates word identification from word fragments. For example, in (Nooteboom 1981) Dutch speakers were presented two different types of Dutch word fragments: word-initial fragments and word-final fragments. Both types of fragment shared the characteristic that they uniquely identified their source words.19 Interestingly, even in this configuration the initial word part served as a much better cue to correct word identification than the final part. Thus, subjects identified the target word correctly in 95% of all pertinent responses20. By contrast, if given the final fragment of the target word as a stimulus, the proportion of correct identification was only 60%. The results of

---

18 Other commonly remembered parts are the stress pattern and final word parts.
19 In fact, the source words for the initial and the final fragments were the same words. The set of stimuli comprised 15 Dutch words, which are, by accident, uniquely identifiable from both their initial and their final part.
20 i.e. those in which the stimulus was heard correctly and in which no other problems intervened, cf. Nooteboom 1981: 414f. for discussion.
Nooteboom's study thus not only show that the initial part of a word plays a key role in word recognition; it also demonstrates that the initial portion of a word often suffices to lead to correct identification (95% of Nooteboom's stimulus words). It is, however, important to note that at this point Nooteboom's experiment ceases to be a good simulation of the task that hearers of truncated names and words face when trying to recover the base of a truncated word from the truncation. Thus, on the one hand hearers of truncated words are put at a disadvantage if compared to Nooteboom's subjects in that certainly not all truncated words uniquely identify their bases. On the other hand, however, hearers of truncated words are privileged compared to Nooteboom's subjects in that, in addition to the initial word fragment, they have contextual cues (both discourse context and situational context) at their disposal to aid them in determining the base of a truncation. Whether or not context is able to balance out the fact that truncated words not always uniquely identify their bases is a question that has to my knowledge not been addressed in empirical research so far.

The role of stress-related information in word-recognition is less straightforward than that of initial material. Whereas there is abundant evidence that shows that stress is important in word recognition in many languages, it is not so clear whether what is relevant for word recognition is the stressed syllable itself, or the prominence pattern of the whole word, and how differences between individual stress-based languages with respect to the role of stress in word recognition are to be explained (cf. esp. Cutler & Pasveer 2006 for an overview of the relevant findings). Psycholinguistic research has focussed on investigating to which extent stress information is exploited for word recognition (cf., e.g., Cutler & Carter 1987, Slowiaczek 1990, Soto-Faraco et al. 2001, van Donselaar et al. 2005, Reinisch et al. 2010), and there are a number of recent studies that demonstrate that stress information plays a very important role. For example, (van Donselaar et al. 2005) and (Soto-Faraco et al. 2001) have shown for Dutch and Spanish, respectively, that in priming experiments, primes consisting of the first two syllables of the target words lead to faster reaction times and fewer errors in lexical decision tasks if the fragments used as primes contain the right stress information, than if the fragments have a different stress pattern. The latter even inhibit word access. More recently, (Reinisch et al. 2010) have convincingly shown in an eye-tracking experiment that Dutch listeners indeed use stress information already in very early stages of word recognition to disambiguate between segmentally identical, but prosodically different competitors (e.g. upon hearing the first two syllables [okto] of Dutch Október vs. Óctopus). Their findings suggest that the moment a word fragment becomes available for auditory word recognition, listeners use not only the segmental makeup of that fragment for their word search, but also
the information of whether or not that fragment bears stress. Crucially, relevant stress information seems to reside not only in the perception of the acoustic difference between successive strong and weak syllables, but also in the perception of strong syllables themselves. Thus, on the basis of an acoustic analysis of their stimuli and a correlation of these with the fixation data, (Reinisch et al. 2010) found that perception of a strong syllable lead to a decline of the number of fixations on competitors with a weak syllable in this position. The relevant acoustic parameter for (Reinisch et al. 2010)'s Dutch informants was the duration of the vowel in the syllabic nucleus.

Stress has also been shown to be important in earlier stages of auditory processing than word recognition. Thus, in stress-based languages like English and Dutch, stress is used by listeners as a segmentation cue. Evidence from misperception of word junctures and word-spotting experiments suggests that the set of competing lexical items activated upon hearing words with non-initial main stress includes words starting with the syllable that is stressed in the target word (e.g. Cutler & Norris 1988 et seq. for English, Vroomen et al. 1996 for Dutch).

In sum, we know that lexical stress plays an important role in word recognition in many languages. Likewise, evidence from studies investigating segmentation suggests that in languages such as English and Dutch, the expectation of listeners seems to be that words begin with strong syllables. For the recoverability of bases in truncation in such languages this suggests the following interpretation: For stress-anchored truncatory patterns, base recoverability is more difficult if the base does not bear initial stress than if it does. This is due to speakers' biases in segmentation. However, findings like those of (Reinisch et al. 2010) for Dutch suggest that stress-anchoring for bases with non-initial stress also has advantages over initial-syllable anchoring: The advantage is that this way the truncated form preserves the stressed syllable of the base as a stressed syllable, preserving the stress cue that is relevant for word recognition. This is not always the case if a truncated form anchors to the initial syllable of a base form. If stress in that base is non-initial, a trochaic truncation anchoring to the initial syllable would alter the prosodic cue to be used in recognition.21

To sum up our discussion of anchoring, we see that there are two types of anchoring that can be established as systematic anchoring patterns in a variety of different truncatory patterns across languages: initial anchoring and stress anchoring. We have then seen that the anchoring facts found in truncation do not seem to be accidental: They are the patterns best suited to

---

21 Note that in stress-based languages which have vowel reduction the argument that initial-syllable anchoring alters cues that are important for recognition is even stronger. Compare, for example, regular differences in vowel quality between truncated names and their bases if main-stress in the base is non-initial: e.g. P[æ]tricia > P[æ]tr, [ɔ]lonzo > L[ɔ]n.
preserving cues that are vital for word recognition. Thus, in the patterns investigated here, anchoring systematically facilitates recoverability of base forms. Truncated forms are as transparent as possible, within the limits set by constraints on word structure.

3.3 The homonymy problem

In this section we will look at the question of how strongly recoverability of bases of truncated forms is influenced by the potential of truncation to create homonymous forms. I will again base my discussion on examples from English.

It has often been noted that truncation may lead to a large number of homonymous forms. For example, names like *Alonzo*, *Alfred*, *Alvina*, etc. may all be truncated to become *Al*. Similarly, among clippings, there are at least three homonymous forms *mag*, which are derived from *magazine*, *magnesium*, or *magnet*, respectively. For obvious reasons, then, the potential of truncatory processes to lead to homonymous forms creates problems for base recoverability.

In what follows we will take stock of the distribution of homonymous forms in the truncation corpora used in (Lappe 2005), (Lappe 2007). Table 19 provides an overview of the pertinent name truncation data (cf. Table 1 above for references).

<table>
<thead>
<tr>
<th>N</th>
<th>%</th>
<th>examples:</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 truncated name has 1 base</td>
<td>498</td>
<td>74.3%</td>
</tr>
<tr>
<td>1 truncated name has 2 bases</td>
<td>107</td>
<td>16.0%</td>
</tr>
<tr>
<td>1 truncated name has 3 bases</td>
<td>44</td>
<td>6.6%</td>
</tr>
<tr>
<td>1 truncated name has more than 3 bases</td>
<td>21</td>
<td>3.0%</td>
</tr>
<tr>
<td>Σ</td>
<td>670</td>
<td>100.0%</td>
</tr>
</tbody>
</table>

Table 19. no. of truncations by no. of bases in Phillips 2002

We see that the corpus, which comprises 948 different base-derivative pairs (cf. Table 1 above) contains 670 different truncated names. Of these, 498 (i.e. 74.3%) have only one single base, and, therefore, no homonymous forms, in the corpus. For another 107 truncated names (i.e. 16.0%), two different base forms are attested. The table thus shows that most
truncated names in the corpus have only one base form, and only 3% of all truncated names in the corpus have more than three base forms. The form with most base forms is *Bert*, for which ten bases are attested: *Adelbert, Albert, Alberta, Bertram, Delbert, Egbert, Gilbert, Herbert, Hubert,* and *Roberta*. Note, however, that the distributional facts displayed in Table 19 clearly mark the situation of *Bert* as exceptional.

Table 20 provides the analogous distribution for clippings in the pertinent corpora (cf. Table 1 above for references).

<table>
<thead>
<tr>
<th>N</th>
<th>%</th>
<th>examples:</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 clipping has 1 base:</td>
<td>532</td>
<td>87.4% <em>barb</em>: <em>barbiturate</em></td>
</tr>
<tr>
<td>1 clipping has 2 bases:</td>
<td>63</td>
<td>10.3% <em>mill</em>: <em>million, millimetre</em></td>
</tr>
<tr>
<td>1 clipping has 3 bases:</td>
<td>13</td>
<td>2.1% <em>mag</em>: <em>magazine, magnesium magneto</em></td>
</tr>
<tr>
<td>1 clipping has 5 bases:</td>
<td>1</td>
<td>0.2% <em>mono</em>: <em>monochrome, n, adj, monophonic, n, adj, mononucleosis</em></td>
</tr>
</tbody>
</table>

| Σ         | 609   | 100.0% |

Table 20. no. of truncations by no. of bases in Beale (1989), Spears (1989), OED

All in all, the corpora, which contain 702 different base-derivative pairs (cf. Table 1), comprise 609 different clippings. Of these, 532 (i.e. 87.4%) only have one single base form. We therefore not only see that most clippings in the corpus have only one base form – homonymy also seems to be much less widespread among clippings than among truncated names. Finally, there is only one clipping that has more than three bases (*mono*, for which five bases are attested).

In sum, we see that, at least for English name truncation and clipping, the homonymy problem is in fact smaller than may be expected. Although it undoubtedly exists, homonymy seems to be the exception rather than the rule among truncated forms, a fact that may be related to the conditions under which truncations are actually used in context (cf. below).

Apart from the question of how widespread homonymy is among truncation, a second issue that is related to the homonymy problem is how speakers and hearers cope with the homonymy problem in language use. This problem is related to two wider sets of issues,
namely to (a) how speakers and hearers cope with homonymy in language in general, and to (b) how, specifically, truncations are used in context. Section 3.4 will be devoted to this latter issue, focussing again on English.

3.4 The role of the context

A standard assumption in the morphological literature is that truncations arise in circumstances in which contextual factors are tightly constrained so that base recoverability is possible (cf., e.g., Jespersen 1965repr: 538ff. and Marchand 1960: 363f. on English). These are, specifically, in-group slang, all kinds of specialised language, and, for names, 'the narrow family circle' (Jespersen 1965repr: 540). It is, however, important to note that the use of truncation is by no means confined to such contexts. For English, both Marchand and Jespersen interpret this as a secondary development, where outputs of truncation, which have originated in the tightly knit situational contexts just described, come to be used also outside these circles.

No matter what exactly lexicalisation or conventionalisation paths of truncations look like, it is interesting to investigate how truncations are used when they are used outside the narrow contexts mentioned by Jespersen and Marchand. Here we can observe the use of truncations in situations in which speakers or writers assume that hearers and readers without specialised contextual knowledge will be able to decode them, and, hence, in many cases, that hearers and readers will be able to recover bases of truncated words from their lexicon. Apart from theoretical assumptions made in the literature, there is, to my knowledge, no systematic empirical research that addresses this issue to date. Therefore, this section will be confined to a presentation of preliminary evidence gathered mainly from my own collected materials. The conclusions to be drawn from this discussion are necessarily tentative, leaving it to future research to test them against larger amounts of data.

An interesting test case for the question of how truncations are used in such contexts is the use of clippings, i.e. truncated non-names, in media that are aimed at a wide and, crucially, non-specialised audience. Here we can approximate the lexical resources that hearers and readers have at their disposal for decoding bases using standard dictionaries and electronic corpora.

In what follows we will look at a small selection of quite recent English clippings from the *Time Magazine*. They are provided in (12), together with the immediate sentence context in which they are used in the source.
(12) a sample of clippings from the *Time Magazine* (clippings are underlined)

a. Even more robust than the lowbrow *merch* trade is the market for knockoffs of items worn by the future princess [...] (April 25, 2011, p. 48, in an article about the royal wedding of Prince William and Kate Middleton)

b. Think Tank [a band] is often experimental but never jarring, mixing *synth* and acoustic picking (May 12, 2003, p. 57)

c. the latest news on how your body handles *carbs* vs. fats [...] Should you count calories or *carbs*? (July 07, 2003, p. 50)

d. Last Thursday, when Arnold Schwarzenegger arrived at the county government building in Norwalk, Calif., you could tell with no trouble that he was one of the biggest stars in Hollywood - and not just if you measure *lat* spread. (August 18, 2003, p. 18)

e. His private life, like that of most Delta Force *vets*, is largely hidden... (Nov 03, 2003, p. 27)

With the exception of *merch*, all clippings cited in (12) are attested as lemmas in the *Oxford English Dictionary* (OED). As earliest attestations, the OED cites the dates 1891 (*vet* < *veteran*), 1939 (*lat* < *latissimus dorsi*), 1976 (*synth* < *synthesizer*), and 1981 (*carbs* < *carbohydrates*). For all clippings cited, we may hypothesise that base recoverability does in some sense play a role in decoding the meaning of the sentence. Thus, three of the five clippings are listed by the OED as homonymous clippings for different base words: *carb* is attested as a clipping of *carbohydrate* and *carburettor*, *lat* is a clipping of *latissimus dorsi* and *latrine*, and *vet* is listed as a clipping of *veteran* and *veterinary* (*surgeon*). Interestingly, the meanings of these homonymous pairs are so unrelated that it is hard to think of a context in which homonymous forms could be confused. For at least another three of the five clippings, we have evidence that they have their origin in specialised language in the sense of (Jespersen 1965repr: 538ff.) and (Marchand 1960), which means that it is unlikely that the writers of the articles cited in (12) can safely assume that the clippings are lexicalised for all readers of the *Time Magazine*. *Merch* (12.a) is not even attested in the OED, which is an indication that the clipping is very recent. A Google© search for *merch* on US American websites reveals that *merch* is mainly used to refer to merchandise articles sold, in particular, to young fans of movies, rock 'n roll music and related arts.  

---

22 A typical usage is exemplified in the following quote from a website promoting a rock 'n roll tribute movie to the *Harry Potter* books by J.K. Rowling: 'We hope you'll still find our *merch* to be affordable and apologize for
specialised language in the music and pop culture. The OED lists three attestations for *synth*: Two of them (1976 in the *Liverpool Echo*, 1983 in the *Yellow Advertiser*) are private ads advertising second-hand syntheses, and one is a review of a band's performance in the music magazine *Sounds*, 1977). Finally, the OED marks the clipping *lat*(s) (s.v.) explicitly as a term from bodybuilding language and specifically mentions the attributive use of *lat* in *lat spread* as referring to a bodybuilding pose.

What we see in the examples cited is that the immediate context in which the clipping can be used to identify the base forms of the clippings. Thus, in all but one cases (*synth*, x.b), the syntactic context makes it clear that both the clipping and its base must be a noun. For *synth*, the context is ambiguous, allowing for an analysis as a noun or as an adjective, respectively.

We also see that readers of the sentences are given strong semantic cues in the immediate environment of the clipping that serve to constrain the search for base words. In the cases of *merch trade*, *lat spread*, and *Delta Force vets*, clippings are embedded in a nominal compound where the meaning of the other elements already defines the semantic field within which the base of the truncation is to be found, and, in the cases of *lat* and *vets*, excludes the homonymous competitors of these clippings. The clippings *synth* (12.b) and *carbs* (12.c) are embedded in coordinating constructions, where the clipping is coordinated with an antonym, *acoustic* for *synth* and *fats* and *calories* for *carbs*, respectively. Again, thus, the construction excludes the homonymous competitor for *carbs*, *carburettor*, and constrains the semantics of *carbs* and *synth* to antonyms of *acoustic* and *fat* or *calories*. This, for example, immediately precludes the possibility that any of the many English words or senses of words starting with <synth> and referring to concepts from philosophical (e.g. *synthesis*, *synthetic*, *synthesist*) or chemical (e.g. *synthalin*, *synthase*, *synthetic*, *synthesise*) semantic fields, would act as bases of *synth*. Similarly, it excludes all words starting with <carb> and referring to chemical compounds which are, at least to lay wisdom, not relevant to dieting, such as, for example, *carbon*, *carbon dioxide*, *carbonium*, or *carbonate*.

Apart from providing semantic cues, it may also be the case that the context in which clippings occur facilitates base recovery on a much more simple, straightforward level. Thus, it is known that speakers and listeners keep some sort of statistical record of cooccurrence probabilities of words and constructions in discourse (cf. e.g. Jurafsky 2003 for a summary of evidence and for pertinent references). It is therefore possible that, on a mere formal level, the immediate context in which a (potentially unknown) clipping occurs already influences base recovery on the basis of a probabilistic estimation of the chances that any of the potential...

---

competitors for a base occur in a given formal environment. Methodologically, we can test this hypothesis by approximating cooccurrence probabilities in large electronic corpora that are balanced to reflect a (broadly) representative sample of the contemporary language as it is written or spoken.

As an example, we will apply the procedure to the clipping merch in (12.a). Recall from the discussion above that merch is (a) not attested in the OED and (b) that it has its origins in the special language of a very narrowly defined community. This makes merch a good candidate here, because, of all examples in (12), merch seems to have the lowest degree of lexicalisation. In addition, recall that in our attestation in (12.a) merch occurs in a very tight compound construction, which makes it easy to define the context over which we would want to determine cooccurrence probabilities.

We will first take the OED as a reference point for what potential candidates for bases of the clipping merch could be. A search in the OED for nouns that are in current usage (i.e. non-obsole) starting with the letter sequence <merch> yields 23 hits. Although all 23 words are, with one exception (merchet), members of the same word family, we may still assume that recovery of a base of merch in the given context involves singling out one of several competitors, such as, for example, merchant, merchandise, merchandiser, merchantability, merchanting.

In order to estimate how the immediate context of merch in our example (12.a) provides a cue to its base, we will use the Corpus of Contemporary American English (COCA, 425 million words, BYU interface at http://corpus.byu.edu/coca/) to determine the likelihood that any of the competitors starting with <merch> occurs in the same construction.23 The search targeted words starting with the sequence <merch>, and immediately preceding the noun trade. The first finding is that, like in the OED, the combination merch trade, involving the clipping, is unattested among all 425 million words in the COCA. This supports the assumption that merch is indeed very recent. The second finding is that if we look at the distribution of competitors cooccurring with trade, we see that indeed the given context is very well suited to biasing the base search for merch towards the intended competitor, merchandise. Thus, of the 23 competitors for a base for merch that we found in the OED, only two different words are attested as cooccurring with trade. These are merchandise and merchant, which occur in 72 places in the corpus in total. Of these, 69 attestations (i.e. 95.83%) involve merchandise trade. The facts for merch thus suggest the following: Clippings seem to often occur in contexts in which also their bases are highly frequent. More specifically, their bases are more

23 The choice of the COCA was motivated by the fact that the Time Magazine, from which (12.a) has been taken, is a US American magazine.
frequent in these contexts than the bases of their competitors. Given what we know about the use of cooccurrence probabilities in speech processing, this insight, if verifiable by future research, has strong implications for an interpretation of base recoverability in word clipping.

A third, related type of contextual cue that is likely to play an important role for base recoverability in truncation is frequency of use of the base of the truncated word itself, within the speech community in which the truncated word is used. Thus, for English it has been shown that the meaning of (most) truncatory patterns involves signalling familiarity and closeness with the referent of the base (e.g. Wierzbicka 1984, Quirk et al. 1985, Schneider 2003, Plag 2003: 117, 121, Alber & Arndt-Lappe to appear). We can therefore assume that the base as well as the derivative are highly frequent for speakers using truncated words. Crucially, for these speakers, the base form is presumably more frequent than competing other base forms. To take again our example merch (12.a), we may assume that within the pertinent speech community in which merch is used regularly (young fans of movies, rock 'n roll music and related arts, cf. above), the base merchandise is more frequent than its most serious competitor, merchant. Likewise, for name truncation, it is plausible to assume that among close acquaintances of the person referred to, the base name for a given truncated name will be used more frequently than its potential competitors. For example, among close friends and family of a person named Lucinda whose conventionalised nickname is Cindy, the name Lucinda will be used more frequently than other potential bases of Cindy, such as Cinderella or Cynthia (cf. Table 19). This high relative frequency of the base form will in itself be another factor that enhances recoverability. The reason is that we know that high frequency of a word facilitates lexical access (cf. esp. Segui et al. 1982).

To sum up this section, we see that our findings concerning the role of context in enhancing base recoverability is similar to what we found for anchoring: The conditions under which truncations occur in context are optimal conditions when it comes to facilitating word recognition and lexical access. We also see, however, that much more research is needed to explore the exact interaction between contextual factors and the use of truncation.

Furthermore, on a theoretical level, it remains an open question if and in how far the conditions of use of truncation in context are different from those of other word-formation processes.

4. Conclusion

24 The idea for this argument is due to Ingo Plag, p.c., July 2006.
Using a selection of case studies, this paper investigated structural predictability and base recoverability in name truncation and clipping. Structural predictability was investigated on three levels: word structure, anchoring, and segmental stability. For word structure, we looked at three independently collected samples of different sizes from English, Italian, and German. The results were surprisingly parallel: the overwhelming majority of forms (usually > 90% in a sample) fall into a well-defined set of systematic patterns. Furthermore, whereas there are cross-linguistic similarities between many of these patterns, we also saw that not all patterns are represented in all languages, and that, within languages, the answer to the question which patterns are used for name truncation and which are used for clippings is clearly language-specific. For anchoring, we investigated the same samples and saw that, like word structure, anchoring is extremely systematic, with again > 90% of all data in each sample falling into one of three patterns: initial anchoring, main-stress anchoring, and a pattern which retains initial and main-stressed material. Variability between anchoring options is, like variability in word structure, determined by the truncatory pattern. For segmental stability we discussed consonant stability in English truncated names and clippings. There we saw that the overwhelming majority of name truncations and clippings faithfully copies pertinent consonants from their bases. Among those truncations that do not, there is one pattern of systematic changes, involving substitution of the dental fricatives. We also saw, however, that avoidance of the dental fricative is a property that is confined to the cophonology of name truncation, and does not occur in clipping. In general, then, we found that outputs of truncatory patterns are highly predictable, under the proviso that predictability does not preclude the existence of systematic variability. The latter is a general characteristic of truncation, but that this systematic variability only occurs within the limits set by the particular cophonology of the truncatory pattern (i.e. name truncation and clipping, respectively, with their suffixed and unsuffixed subpatterns). On the basis of the fact that systematic variability seems to be tied always to specific truncatory patterns (and is, hence, morphologised), I argued in this paper that systematically variable aspects of truncation have a different status in the analysis than aspects that are idiosyncratic to isolated forms. The former are part of the grammar of truncation, the latter are not.

On a theoretical level, then, the findings on structural predictability challenge the view sometimes found in the literature that outputs of truncation are, in general, unpredictable. Furthermore, it suggests that the sharp line that some approaches assume to exist between grammatical word-formation and extra-grammatical processes cannot be upheld (Dressler 2000, Dressler 2005, Ronneberger-Sibold 2010, Ronneberger-Sibold in prep). Instead, the
evidence from name truncation and clipping implies that there is a gradient transition between what we may consider to be core processes of word-formation and word creation. The second part of the paper was concerned with the question of base recoverability, an aspect of morphological processes that is related to transparency. First of all, it was shown that this is a notoriously understudied area, where many of the pertinent issues have not been addressed by empirical research in a systematic fashion. We then discussed three types of evidence that can tell us something about base recoverability in truncation: anchoring patterns, homonymy, and the role of the discourse context.

For anchoring it was shown that, in spite of the fact that outputs of truncation are formally non-compositional, anchoring in truncation systematically facilitates recoverability of base forms. Within the limits set by constraints on the word structure of the truncated form, the truncated form preserves the part of the base which is most important for word-recognition. With respect to the homonymy problem, we saw that, at least within the corpora taken from (Lappe 2005), (Lappe 2007), less homonymy exists among English truncated forms than may be expected. It is an open question why this is so. Potential reasons could be, for example, that the existence of a truncated form blocks the formation of homonymous forms. Similarly, it could also be that truncation preferably occurs with base words which can already be uniquely identified on the basis of a monosyllable. Investigating the role of contextual factors in base recovery, the paper presented evidence that suggests that context plays a very important role for base recovery. We discussed a small sample of attestations of English clippings in context from the Time Magazine. In all examples processing of the sentence involved either disambiguation of homonymous lexicalised clippings or clippings which are unlikely to be lexicalised for the bulk of the intended readership of the relevant article. It was shown that in all examples the context supplied very clear semantic cues in the immediate syntactic environment of the clipping that aided recovery of the intended base among all potential competitors for bases. Using data from the OED and the COCA corpus, we also saw in the case of the recent, non-lexicalised clipping merch, we could even predict the intended base by comparing cooccurrence probabilities in the COCA of semantically similar and plausible competitors for bases. Finally, I have argued that yet another factor that strongly facilitates base recovery is the relative frequency of occurrence of a base as compared to that of its potential competitors, in the language used by the speech community in which the truncation has been conventionalised. Again the argument is based on what we know from psycholinguistic research on word recognition, where frequency has been shown to play a similar role.
In sum, the discussion has shown that, due to the way it is often defined in the literature (involving a compositional element), the notion of transparency is difficult to apply to truncation. It thus proved helpful to reformulate the question of whether or not truncations are transparent, in terms of base recoverability, and to relate issues of base recoverability to the psycholinguistic issues of word recognition and lexical access. Truncations are thus different from other, compositional morphological processes in that establishing a relation between a truncated form and a base is, essentially, word recognition on the basis of a fragment.\(^{25}\) The evidence discussed then clearly suggests that, in spite of all the problems that truncation theoretically poses for base recovery, truncations are actually formed (in terms of anchoring and homonymy avoidance) and used (in terms of contextual cues) in such a way as to create ideal conditions for word recognition.

\(^{25}\) The truncation-specific semantics may be assumed to fall out naturally of the fact that the truncated form is only a fragment, and not the full form.
References


Libben, Gary, Martha Gibson, Yeo B. Yoon & Dominiek Sandra, 2003, 'Compound fracture: The role of semantic transparency and morphological headedness', in Brain and Language 84:1, 50–64.


Poynton, Cate, 1989, 'Terms of Address in Australian English', in *Australian English - The Language of a New Society*, Peter Blair & David Blair (ed.), St. Lucia, Queensland, 55-69.


Schneider, Klaus-Peter, 2003, *Diminutives in English*. Tübingen.


Sundén, Karl, 1904, Contributions to the Study of Elliptical Words in Modern English. Upsala.


