LEXICALIZATION AND INSTITUTIONALIZATION

PETER HOHENHAUS

1. INTRODUCTION

Both terms in the main title of this chapter usually refer to what can happen to words during the course of their 'life' *after* they have been formed. Despite the implication of 'after' and 'life', this does not mean that they are relevant *only* as purely diachronic concepts. Even though many usages of *lexicalization* do in fact primarily revolve around diachronic phenomena, such as formal *demotivation* or *semantic idiomatization*, the more interesting questions about lexicalization and *institutionalization* relating to word-formation theories involve important synchronic issues. These include: the nature of the lexicon, the extent to which complex forms need or need not/must not be listed in it, what facilitates/prevents the listing of new complex formations, and the balance of idealization vs. sociological and psychological reality.

2. LEXICALIZATION

Of the two, this is the more widespread term. However, there isn't complete consensus about (the scope of) its meaning. In the following I outline the main readings of the term.[1]

2.1 Lexicalization in a diachronic sense

The least controversial and fairly well understood aspect of lexicalization is certainly the fact that vocabulary items tend to change over time. While, as Welte (1996: 79) points out, the terminology in early descriptions in the 19th and early 20th century varied even more than it does today, the phenomena as such are hardly recent discoveries:

Formal changes that complex words can undergo include phonological phenomena such as vowel reductions as in [mæn] > [mən], e.g. in *policeman*. A step

---

[1] There are usages of the term that go beyond the focus of this chapter. In NLP ('Natural Language Processing', in Computer Linguistics), for instance, 'lexicalization' is used in a basically onomasiological sense, as the filling with lexical material of more abstract conceptual structures at intermediate stages of processing. Lipka (2002: 111) mentions similar uses of the term within Generative Semantics. In some diachronic language studies, 'lexicalization' is used as a kind of antonym to 'grammaticalization', although, as Lehmann (2002) has pointed out, this distinction is not quite so simple, namely in that grammaticalization presupposes prior lexicalization of the item/construction in question.

up from such minor changes are cases such as *breakfast* [bɹeik fɑːst] > [brekfəst], or changes resulting from stress shift, e.g. *admire* [ədˈmaiə] – *admirable* [ˈædmərəbl].

A textbook example of extreme formal lexicalization is *fo'c'sle*, from *forecastle*, pronounced [fəʊksl], 'sleeping quarters under the front deck of a ship', where the abbreviated spelling reflecting the pronunciation practically makes it an unanalysable simplex even in the written form (cf. Lipka 1977, 2002: 113). Curiously, though, such lexicalized forms sometimes become remotivated in so-called spelling pronunciations such as [weistkəʊt] instead of the older pronunciation of *waistcoat* as [weskət] (cf. Bauer 1983: 52f).

Formal lexicalization phenomena are not limited to phonological ones. Welte (1996: 80) mentions e.g. the dropping of the otherwise obligatory plural morpheme of words such as *trousers* or *pyjamas* when part of a compound: *trouserpocket*, *pyjama top*, while in Swedish we find a reverse phenomenon: the plural of *en liten stad* 'a small town' is *små städer* – and the latter also has to be selected in compounds (in the singular!): *en småstad* (cf. German *eine Kleinstadt*) or *småkaka* 'small cake', 'biscuit'. Furthermore, in lexicalized words we find structures that would be ungrammatical as a freely constructed syntagma: (*to score an*) *own-goal*, but e.g. **to have an own house*. (Cf. also lexicalized *forget-me-not* vs. **forget me not – don't forget me*.)[2]

Many discussions of lexicalization focus on demotivation and/or loss or addition of semantic features – and alternative terms have been used by different scholars, including *fossilization* (Lyons 1977: 547), *petrification* (Leech 1974: 226), and in particular *idiomatization*.

Kastovsky (1982a) makes the additional distinction between *systematic lexicalization*, such as in the regular addition of very general features such as [+ PROFESSIONAL] in derivations by means of *-er* (*lecturer*, *reporter*, *writer*), and non-systematic, i.e. truly idiomatic semantic lexicalization.

Aronoff (1976: 19, 43) frequently uses the term *semantic drift* here. A famous example of his is *transmission* not as a regular action nominalization but as the technical term for a part of a car – namely the one that transmits the power of the engine to the wheels.[3] So this semantic specialization can be described as addition of semantic information (regarding object, source and goal of the verb).

Bauer (1983: 55ff) finds accounts of semantic lexicalization as additional semantic information problematic, partly because criteria can be mixed, e.g. Lipka's (1977) distinction between *language-internal changes* as opposed to *extralinguistic changes* in cultural background. Examples for the former are *mincemeat* or *sweetmeat* (which *are* suitable for vegetarians) where the second element goes back to an older broader meaning, 'food in general', but which has later narrowed to its

---

[2] Bauer (1983) makes a few further distinctions regarding, for instance, the so-called combining forms (German 'Fugen') in Danish and German compounds, the isolation of learned roots such as in English *edible*, or that of 'syntactic' patterns such as 'verb-object', now unproductive, in compounds like *pickpocket*.

[3] I.e. the verbal part of the formation *is* motivated; some morphologists would analyse this as a zero-headed construction

current meaning. An example for the latter is *watchmaker* – a profession that these days typically no longer includes the making of watches, only repairs.

Bauer (1983: 57f) also sees a problem in specifying what amount of semantic information may be lexicalized. One of his examples is the exocentric compound *redskin*. This has two established meanings, although not all speakers may be familiar with both: a) a person (now a rather politically incorrect term) and b) a type of potato. This in itself shouldn't be such a problem as it is far from unusual for one form to have more than one meaning and thus more than one lexical entry (polysemy and homonymy are all over the place in the lexicon anyway). However, Bauer (1983: 58) also claims that the relevant additional semantic information is rather down to context than to specializing lexicalization, since *redskin* can be reinterpreted with ease as in the following context: [4]

(1)   Granny Smith was rude to all lesser breeds, but particularly to Mr Mackintosh. "You redskins", she would sniff, "You're all alike: no firmness of character."

What Bauer fails to acknowledge here, however, is that this instance of word-play – a case of a "nonce-use" (cf. Hohenhaus 1996: 133ff) – relies on the deviation from the lexicalized, conventionally fixed sense of *redskin* (presumably rather the 'person' reading), otherwise we would not recognize it as such. Consider also the more recent example *warhead*. Normally this is lexicalized as 'the explosive front end of a missile'. In news coverage of the 2003/2004 demonstrations against the Iraq war, however, banners could be seen that had a picture of George W. Bush or Tony Blair next to this word, clearly prompting a different, nonce reinterpretation.

Addition or loss[5] of semantic information through lexicalization, and demotivation, is often referred to as idiomatization, to cover the general aspect of lack of compositionality, which such lexicalized words indeed share with idioms proper such as *to be over the moon* 'to be very happy'.

It is important to note, however, that idiomatization is only one aspect of lexicalization, which is why the two terms should not be used interchangeably (as is sometimes the case). Rather 'lexicalization' has to be regarded as the cover term for a range of phenomena, semantic and non-semantic. Bauer (1983: 49) also emphasizes that "opacity is not a necessary pre-requisite for lexicalization" since "[s]ome lexicalized forms […] may remain perfectly transparent", e.g. *warmth* – which must be considered lexicalized because "the suffix *-th* cannot be added synchronically to an adjective to provide a noun".

Finally, the various subtypes of diachronic lexicalization phenomena are not neatly separated issues but frequently overlap, e.g. *holiday* is not only phonologically different from *holy day* but also semantically specialized: of the second constituent only a feature like [PERIOD OF TIME] remains, while an "inferential feature […] of the first constituent, namely [NO WORK] has become an

---

[4] The exact source is not given, but it is linked to an advertisement for an apple orchard.

[5] An oft-quoted extreme example of loss of features is *understand*, where both constituents have practically lost all the semantic components of *under* and *stand* as free forms.

obligatory feature of the whole lexeme" (Lipka 2002: 114). A similar example from derivation is *infamous*, phonologically lexicalized as [ˈɪnfəməs] and semantically lexicalized in that it doesn't mean 'not famous' but 'famous in a negative sense'.

Lexicalization in the diachronic sense is clearly a gradual affair, both diachronically (proceeding in successive stages over time) and in the synchronic description of the results of such processes (cf. Lipka 2002: 113). Thus we have a cline ranging from complete formal and semantic opacity, with the results becoming similar in status to unanalysable simplex words (e.g. *gospel*), via partial idiomatization/demotivation, and minor vowel reductions, to even fully transparent forms such as *warmth* (despite the synchronically non-productive suffix).

But why do words undergo diachronic lexicalization in the first place? One obvious and immediately intuitively plausible explanation is frequency of use. Lipka (1981, 2002: 111) has repeatedly stated this as the main reason. However, Lipka (1977: 155,161) also points out that the phenomenon of *hypostatization*[6] can be a reason for lexicalization initially.[7] Hypostatization is a side-effect of the naming-function of word-formation, whereby the existence of a word seems to imply for speakers the existence in the real world of a single corresponding 'thing' or clearly delimited concept.[8] Indeed, this could help trigger semantic specialization: making one out of several readings of a potentially ambiguous formation become fixed, the word thus becoming a sign in its own right, losing its character of a syntagma – which is the general default characteristic of lexicalization in the diachronic sense (Lipka 1977: 156).

2.2 Lexicalization in a synchronic sense: listing/listedness

The term is itself potentially ambiguous, as *lexicalize* + *-ation*, can be interpreted as a) an action nominalization or b) a result nominalization. In a synchronic sense, then, the term would correspond to *the process of listing* (the entering of a word in the lexicon) in a), while b) corresponds to (the state of) *listedness*, i.e. the property of a word of having a lexical entry in the language.

The synchronic relevance of lexicalization in sense b) has frequently been pointed out on various grounds, e.g. by Mark Aronoff in connection with the phenomenon of blocking (e.g. of *stealer* by *thief*). He uses it as an argument to refute Di Sciullo & Williams' (1987) claim that the lexicon was only about the

---

[6] This term is originally taken from Ernst Leisi, but Lipka uses it in a wider sense.

[7] Another cautionary point can be made here: the general assumption that lexemes only acquire 'idiosyncrasies' through *diachronic drift*, is not necessarily quite such a straightforward truism. Herbermann (1981: 334) contests this assumption, claiming that some idiosyncrasies are typically there from the start as part of the coining process, i.e. before an item is listed, so that if it is listed it will have to be listed complete with these features already being present.

[8] In fictional contexts, the hypostatization effect can even be exploited to create 'reality' in a fictional setting, say, in a science fiction novel, where various non-existent objects are named (cf. Hohenhaus 1996: 319ff). Here, of course, the hypostatization effect does not normally trigger proper lexicalization *outside* the fictional context (although a few such words have made it into common vocabulary through the fame of their sources, despite the non-existence of their denotata in the real world, e.g. *time-machine*, *warp speed*, *beam me up* – cf. sections 4.4 and 4.5).

'lawless' (and thus 'boring') and their strong hypothesis that listedness is thus totally irrelevant to word-formation theory. In this context, Aronoff (1988: 767) stresses the asymmetry of the blocking relation "in that the blocking item is in the lexicon" (and the blocked item isn't and couldn't be), i.e. without listedness the phenomenon of blocking cannot be captured, irrespective of whether the listeme is in any way idiosyncratic or not. (Cf. also the example of lexicalized but fully analysable *warmth* above.)

Lexicalization in the sense of listing/listedness overlaps with 'institutionalization' and will be discussed further in conjunction with that concept in section 3.

2.3 The lexicon and theories of word-formation

While it is a fairly uncontroversial commonly held belief that unpredictable idiosyncrasies resulting from lexicalization in the diachronic sense cannot and should not be generated by general rules, but have to be captured piecemeal by lexical entries, the wider role of lexicalization in the synchronic sense and the nature of the lexicon are issues of acrimonious debate. Proposals in word-formation theory generally fall somewhere between two extreme poles; a): a maximally rich lexicon, in the sense of: *all* established/existing/etc. words are listed in the lexicon – versus b): a minimal lexicon listing as little as possible. The former has been dubbed the *full-entry model*, the latter the *impoverished-entry model* by Jackendoff (1975).
The assumption that only as little as possible should 'be in' the lexicon, namely only what is so idiosyncratic that it cannot be captured by any rules expressing general regularities, is a view that is especially commonplace within 'lexicalist' and 'word-syntactic' generative frameworks such as Lieber's (1980), (1992), Selkirk's (1982) or Di Sciullo & Williams' (1987) – see Scalise and Guevara in this volume. This assumption is usually argued for on the grounds of 'conceptual simplicity' of the overall theoretical model, but is also a theoretical necessity for any morpheme-based model (cf. Aronoff 1988: 768). Occasionally an argument in favour of the impoverished-entry model is put forward which is of a psycholinguistic nature, namely that listing non-idiosyncratic formations alongside idiosyncratic items "would only encumber the speaker's memory" (Selkirk 1982: 127). While the former is indeed largely a theory-internal assumption, ultimately down to "choices [made] in terms of the priorities of the linguist" (Bauer 1983: 200), the psycholinguistic argument has to be judged against actual psychological evidence (see below).

However, the central argument for an impoverished-entry model has also always been contested, e.g. by Jackendoff (1997: 124), who points out that its internal logic is less convincing than it may seem: "A fairly standard assumption amongst linguists is that the impoverished entry theory is ultimately correct. […] However, […] although 'conceptual necessity' requires that the lexicon encode what is not predictable, it does not require that the lexicon encode *only* what is not predictable."

Word-based theories (in the wake of Aronoff 1976) should be expected to lean more towards a full-entry model. Aronoff himself remained somewhat undecided on

this theoretical issue (cf. Hohenhaus 1996: 159f); however, he offers a practical argument, namely that a dictionary "is the closest we can come to the lexicon of a native speaker's language" (Aronoff 1976: 116). Herbermann (1981: e.g. 179) also argues that dictionaries have always listed more than just the idiosyncratic minimum. Still, 'large' dictionaries (such as the OED) certainly list a good deal more than can be expected to reflect an average individual's *mental lexicon*.[9] On the other hand, as Bauer (2001: 35f) points out, dictionaries a) represent the *norm*, i.e. the established words of a speech community, rather than an individual's mental lexicon, and b) any reference work is necessarily outdated, lagging behind the rate of new formations. So reference to lexicographic lists of lexical entries can at best serve as a working hypothesis (cf. also Hohenhaus 1996: 362). It does, however, seem to cast some doubt on extremely minimalist (morpheme-based) impoverished entry models.

Such models become particularly doubtful, as Jackendoff (1997: 124) points out, if "we are talking about the mental lexicon, not some mathematically optimal abstract object, so psycholinguistic evidence is absolutely relevant." Indeed, the view that a full-entry model is to be preferred is typically based on the adequacy condition of psychological reality (cf. also Meys 1985, Hohenhaus 1998).

Integrating a refined version of the full-entry model, Jackendoff (1997) develops a highly intriguing, heavily psycholinguistically informed theory of *tripartite lexical licensing* (instead of late 'lexical insertion'). It aims to model parallel processing of phonological, syntactic and conceptual structure, linked by correspondence rules (for interfacing). The 'lexicon' is thus conceived of not as a simple list, but as "a collection of stored associations among fragments of disparate representations" (Jackendoff 1997: 108), which is in line with what is known about other mental processes/representations, only that "the lexicon may be unique in its size and utter arbitrariness" (compared e.g. to visual representations). In my view, Jackendoff's is one of the currently most promising-looking approaches unifying lexicalist generative theory and a broader psychological picture. We will return to psycholinguistic issues, as relevant to our topic, in section 4.3.

Štekauer (2000: 3) takes an onomasiological approach rather than a psycholinguistic one, but by separating a word-formation component from a lexical component (and making them contingent on the speech community and its naming needs as the starting point) he also manages to allow full-entry listing, while still assuming that word-formation as such is 100% regular and productive (see Štekauer, this volume).

A final point that can be made here is that without assuming a full-entry model it would be difficult for a theory to capture certain crucial conceptional distinctions, notably those between *existing words* vs. *possible words* vs. *nonce-formations* – see sections 3.2 and 4.1.

---

[9] While Pinker (1994: 150f) in this context rather flippantly suggests that the number of entries in a commercial dictionary may be deliberately inflated due to publishers' marketing considerations, he still emphasizes: "[t]he brain seems to be reserving an especially capacious storage space […] for the mental dictionary."

3. INSTITUTIONALIZATION

While 'lexicalization' (in one sense or the other) forms part of the terminology of most scholars concerned with word-formation, the term 'institutionalization' is still less commonly employed, although it has gained a certain currency over the last 25-30 years. Broadly speaking, it refers to the stage in the life of a word at (or from) the transitional point between the status of ex-nonce-formation-turned-neologism (cf. section 4.1) and that of generally available vocabulary item, i.e. a formation that is listed but not (necessarily) lexicalized in the diachronic sense yet.

3.1 Terminology

Over two decades ago Bauer (1983: 45) lamented "very confused terminology" in this area. In an attempt to alleviate this confusion, Bauer (1983: 50) suggested we use 'established' as a hyperonym for lexicalization in the diachronic sense, to which he restricts the use of that term, and institutionalization, thus keeping the latter two terms strictly distinct. Laudable as this clarification may have been, it has not become universally shared (we have to bear in mind that we are dealing with notational terms here[10]). Neither is the restriction of 'lexicalization' to its diachronic sense generally accepted (cf. section 2.2), nor is that of 'institutionalization' to fully transparent lexemes only. Often 'institutionalized' and 'established' are used synonymously.

Lipka (2002: 112) adds a different facet of distinction when he says that he adopts the term 'institutionalization' to "stress the sociolinguistic aspects of words" (see next section). By implication, then, a synchronic sense of 'lexicalization' can be retained, but with a focus on lexicological, theoretical aspects – while 'established' may still serve as a cover term for both.

Both Bauer (1983) and Lipka (2002) also employ another (competing) set of terms, introduced by Meys (1975, 1985), namely the dichotomy 'type-familiar vs. item-familiar'. As part of a plea for a full-entry model, this places the emphasis on the individual psycholinguistic aspects involved: a formation is *item-familiar* to a speaker if it is recognized as a particular lexeme, i.e. as one which is not new but already part of the speaker's lexicon, whereas new (regular) formations are only recognized as *type-familiar* but not as listemes in his/her individual mental lexicon (and are therefore analysed, i.e. morphologically processed).

The terminological confusion does not end at the competing definitions already mentioned. Various further terms remain in use, e.g.: 'existing', 'actual' (vs. 'possible'), 'occurring', 'received', 'in use', 'coinage', etc. – all more or less vaguely equivalent to 'item-familiar', 'institutionalized' or 'established'. Whatever notational terms one finally settles for, it is important to relate these to several perspectives.

---

[10] Cf. Lipka 2002, who makes this point frequently, also with regard to various other terms involved, i.e. notational terms do not have a single 'correct' definition, but "can be defined differently in different frameworks" (Lipka 2002: 13).

3.2 Ideal and real speakers and the speech community

Bauer (2001: 34) emphasizes that a category of *existing words* is crucial for a notion of productivity "concerned with the potentiality of new formations" – in order to know what would be new, we have to know what is not new. But he also observes that this entails severe theoretical problems, in particular the question: "existing for whom or what?"

One could argue that an *ideal speaker*'s lexicon contains *all* words of the language in the broadest sense, i.e. also including all potential words since, according to Bauer (2001: 35), it is "unclear how such a construct [as 'the ideal speaker'] could be unfamiliar with a possible word". This is why he rejects the notion of "existing word" as 'existing for the ideal speaker'. I find that rejection a bit too sweeping. How the ideal speaker's mental lexicon is conceived of rather seems to depend on the theory making that idealization. Thus both a minimal impoverished-entry model and a full-entry model are ultimately compatible with the notion of an ideal speaker. For the former, the ideal speaker does not list anything that is predictable by rule, for the latter the ideal speaker memorizes all words, including compositional ones, that represent the permanent vocabulary of a language at any given (i.e. equally idealized) point in time. What is more important is to ask: 'existing/possible *as* what?' – are we talking about potential vs. actual listemes, i.e. the lexicon, or about the output of word-formation processes, independent of whether or not these later enter the lexicon as well? These two perspectives should be kept separate – see section 4.2.

But what about real speakers in real speech communities? The vocabularies of individuals, their lexical idiolects, will vary quite considerably (beyond a certain common core), which is one of the reasons why Bauer (2001: 34), like "most of linguistic theory" (but unlike most psychological studies), rejects "the knowledge of the individual as an irrelevancy."

Lipka (2002: 112), drawing on earlier work by Eugenio Coseriu (1967),[11] suggests that we adopt the notion of *norm* as an intermediate level between *langue* and *parole* – the former of these Saussurian concepts being understood as the 'system' of language, the latter as its individual concrete realizations. 'Norm', on the other hand, is understood as the "collective realization of the language system" (Lipka 2002: 112), and it is this that accounts for, for instance, established choices between systematic alternatives such as *to nationalize* rather than **to nationalify/to national*, or 'habitual disambiguation' of e.g. *sleeping pill* and *headache pill* ('for' vs. 'against').

What is obviously crucial for successful communication is shared vocabulary in speech communities – the latter deliberately put in the plural here, because what is and what isn't shared vocabulary will vary according to different groups of interlocutors which can be collectively taken to constitute speech communities (I assume 'community' to imply a good degree of stability, not just any grouping of speakers). The smaller the speech community, the more it can differ from the

---

[11] Cf. also Kastovsky (1982a: 33, 205).

generally shared vocabulary,[12] along various parameters, on a scale of *idiolect – dialect – language*, cf.. Lipka (2002: 22), where 'dialect' is understood in the broadest sense of the language of any subclass of a speech community.

The smallest setting of a speech community, the subclass just above the idiolect, is that of a couple. Here, intimacy can foster extreme idiosyncrasies – however, due to that very intimacy of such a setting, robust empirical data are hard to obtain. Only very occasionally do such examples surface outside their intimate domain (but readers are invited to use introspection here). A few anecdotal cases are described in Hohenhaus (2005b), e.g. the highly idiosyncratic 'back-formation' of a singular **shoop* from *sheep* – originally a deliberate jocular deviation, which did however become established in the couple's micro-dialect.

The next larger 'community' will be that of the family or other such more or less stable small group (close work colleagues, band members, small teams of explorers on an expedition, etc., etc.). Heringer (1984: 9) mentions the phenomenon of *episodic compounds* for such small groups – a potential example he constructs is German *Mäusebibel* 'mice bible', which is useable by family members who all know about a past incident in which a bible showing teeth marks of mice (who had apparently nibbled at it) was found by the family in a barn. It is thus only on the basis of the common episodic knowledge that the compound can be institutionalized in that meaning within this family's small-group dialect.

At the next higher level lie the special vocabularies of technical jargon, slang etc. – our own context here provides plenty of examples. Acronyms usually make this particularly clear: 'lay' people outside linguistics would hardly be able to decode *NP*, *LFG*, *GB*, *HPSG*, or *OT* (more than one meaning here!). Likewise, acronyms such as *ADSR*, *VCA* or *LFO* (from synthesizer technology) will probably leave most of the readers of this book baffled. Many of the emerging conventions, including more or less institutionalized acronyms, in the new 'e-language varieties' such as Internet chat pose similar decoding difficulties to outsiders or 'newbies', e.g. *lol*, *o4u*, *cul8r*, *cfv*, *imho*, etc. (cf. section 4.4).

Of course, if we want to approach any degree of linguistic generality we have to look at larger settings, ultimately at *the* speech community of *a language* at large, i.e. all its speakers. As the edges of this concept are necessarily fuzzy, this is naturally an idealized entity itself.

At any of these levels, however, *genericness* – based on shared knowledge – seems to be of crucial importance (cf. Heringer 1984, Hohenhaus 1998). The more generic the meaning-form pairing in a word-formation is, the wider its institutionalization potential. The less generic, more context-dependent or individually episodic, the narrower the institutionalization range (cf. section 4.2.).

In short, what institutionalization actually amounts to hinges on what sort of 'institution' we are talking about, ranging from couples and micro-group settings via intermediately-sized groups of jargon speakers to 'the' speech community at large. But such is the sociolinguistic nature of the concept. It is precisely this which makes

---

[12] Cf. McWorter (2001) who repeatedly stresses that the same principle applies generally to languages in a wider diachronic sense: the smaller and more isolated a speech community is, the more 'exotic' features its language tends to develop.

the term a useful addition to our other terminology. It also means that ideally we should always use the term together with its specification 'institutionalized in *X*, *Y* or *Z*'.

3.3 De-Institutionalization: the end of a word's life

We also have to mention briefly that the results of lexicalization and institutionalization are not necessarily permanent states. Obviously enough, words do not just enter a language, they can also 'exit', become obsolete, die out. Lipka (2002: 53) predicts such a fate for e.g. *millennium bug* (correctly, it can be assumed, esp. given that it caused so much less damage than was feared before the turn of the millennium). However, because words, once institutionalized in written varieties, leave a permanent record, they may not completely vanish into oblivion. This may have a distorting effect especially for corpus-based studies, unless frequency is overtly linked with specific time-spans.

Consider another pair of examples of acronyms:[13] *WMD* vs. *IDS*. The former is probably one of the highest frequency acronyms in politics of recent years[14] (although one hopes it will eventually drop out of use together with its referents, but for the foreseeable future this is firmly institutionalized). The latter, on the other hand, was used in a British context[15] for a few years with considerable frequency (esp. in the tabloids) to refer to the leader of the Conservative Party *Ian Duncan Smith*. Since his resignation, the acronym has also notably decreased in frequency and is likely to slowly fall out of use altogether.[16]

The fact that words become obsolete is of course not a new observation. However, it is often assumed that the (English) vocabulary is constantly enjoying 'growth' (cf. e.g. Hughes 2000). While this is certainly true for the *records* that lexicographers keep accumulating, it is unlikely that the mental lexicon, capacious as it may be, can enjoy similarly endless 'growth'. *De-lexicalization*, shrinking or deleting of lexical entries in the mind must therefore also be an important psycholinguistic factor.

---

[13] Lipka (2002: xviii,110, 146) rightly stresses the importance of acronymy as a powerfully productive means of creating new lexemes, even though it poses problems for many word-formation theories in that its morphological status is unclear, since the process is at best only partly rule-governed.

[14] Standing for *Weapons of Mass Destruction* – a Google Internet search on 28 June 2004 returned 768,000 hits in English, a search restricted to sites under a year old still returned 705,000 hits, corroborating the assumption of highly increased current frequency. Sampling and refined searches within results confirm that the majority did indeed include this very acronym – although there were also a few in which it stood for something else, including, curiously, the name of the organization *World Movement for Democracy*.

[15] Of course, for Germanist linguists, the same acronym still stands for the renowned *Institut für Deutsche Sprache*.

[16] It is still to be found, of course. For instance, it was dug out in June 2004 in a BBC Television programme comparing the results of the Tories under their respective leaders in the last few European elections. Empirically assessing the frequency of this word is difficult, a) because the height of its currency and its subsequent decline are too recent to be covered by the available standard corpora, and b) web-searches reveal countless other readings for the letters *IDS*, so that even a time-span-refined search query would require disproportionate filtering efforts.

4. PROBLEMS

As we have already seen, there is considerable disagreement about the terminology and the concepts revolving around the issues under consideration here. Partly, these derive from different theoretical priorities and decisions, but they can also be related to particular empirical and further conceptual problems, whose handling must affect such decisions.

4.1 Nonce-formations and neologisms

The terminological confusion already noted above is perhaps at its worst here. Frequently, the two terms in the title of this section are used almost synonymously.[17] Sometimes 'nonce-formation' is restricted to linguistically irrelevant, quirky stylistic 'novelties',[18] sometimes it is seen as fully representative of the system of word-formation defining 'possible words'.

Along the scale of (tacitly or overtly) suggested meanings for *nonce-formation* I prefer a compromise, similar to Bauer's (1983: 45), covering both perfectly regular outputs of productive rules as well as stylistically (or otherwise) more marked, creative, even deviant 'playful' formations.

Accordingly, in Hohenhaus (1996) I proposed a scalar definition along a set of criteria, including context-dependency and various types of deviation. This is *not* to say that I consider the most deviant, highly context-dependent nonce-formations to be the most typical ones in a quantitative sense (as 'the type most frequently encountered'), *nor* that they should be treated in the same manner as well-formed formations in a theory of word-formation. On the contrary: those nonce-formations that display the most features are in fact the rarest; and deliberate deviation of course has to be separated from regular outputs of rules. The latter distinction is these days often associated with a difference between *productivity* vs. *creativity* – cf. Bauer (2001: 62ff). An example (from *Time* Magazine 08/10/1990, p.90) appears in the following context:

> (2)   It's an oid-y world out there. Tabloids run factoids about humanoids on steroids. In a world gone synthetic, why should movies offer something as organic as a hero? Welcome, then, to the age of the heroid.

---

[17] As far as the sheer number of competing terms equivalent to Engl. 'nonce-formation' is concerned, the situation is even worse in German linguistics. But *Ad-hoc-Bildung*, *Augenblicksbildung* and *Okkasionalismus* seem to be the most frequent choices (cf. Hohenhaus 1996: 17-20)

[18] E.g. Lieber (1988: 206) seems to have such a notion of 'nonce' when she remarks that some productively formed items "do not sound like nonce forms" – this narrow concept of 'nonce' would equate it with 'creative coinings' that are perceived as 'odd'. Even more extreme uses of 'nonce word' can be encountered in psycholinguistic papers (cf. e.g. <www.speech.psychol.ucl.ac.uk/ sr.savagelieven.final.doc>) where it is equivalent to 'non-word' or 'nonsense word'. This is of course most unfortunate from our perspective, but we can only note that we have to remain aware of the great variation in notational terminology here.

Unlike the formation *heroid*, which the text is commenting on (overtly as an innovation), the jocular *oid-y* lacks a proper base: suffix + suffix is not a regularly available morphological pattern in English. Of course, the formation makes *sense* – the semantic function of -*y*-suffixation does apply here: 'characterized by/full of X(es)', here in a meta-communicative variety of that function. Morphologically, however, the formation is rather the result of so-called *rule-changing creativity*. (*Heroid*, in contrast, *can* be taken to represent regular nonce word-formation.)

Similarly, Bauer (2001: 206) mentions the formation of *greenth* by Walpole some 150 years after -*th*-suffixation ceased to be a productive (available) pattern, and he comments that such individual, irregular "innovations are viewed as creating their effect precisely because they are not standardly regular morphology."

However, such creatively deviant formations are comparatively rare (cf. the table in Hohenhaus 1996, appendix II). Far more often encountered are non-deviant but context-dependent nonce-formations, such as the famous example *apple-juice seat* used by Downing (1977) – cf. section 4.2.

The one feature that applies to all nonce-formations, i.e. the necessary (but not necessarily sufficient) condition for 'nonce-ness' as such, is that the formation is 'new' – more precisely: 'new' in a psycholinguistic sense, i.e. formed actively (by whatever means) by a speaker – as opposed to retrieved ready-made from his/her storage of already existing listemes in the lexicon.

This sets nonce-formations apart from neologisms. Neologisms are not new in the absolute sense that nonce-formations are. Rather, the status of neologism is the *next* stage in the life of a word, namely when it begins to be recognized as item-familiar and catches on in the usage of other speakers. Neologisms are thus only new in a relative sense, diachronically, from the point of view of the lexicon. They should therefore rather be described as 'young listemes'.[19]

The problem posed by nonce-formations and neologisms for the concepts of lexicalization and institutionalization are thus linked to the ones they pose for the concept of 'possible word'. Nonce-formations are somewhat 'in between' actual words and possible words: once attested, i.e. having (had) physical reality, they are clearly not (or no longer) merely possible, but nor do they 'exist' in the sense of being part of the lexicon – which is the usual understanding of the notion of 'actual word'. In fact, their existence is typically maximally short-lived: limited to a single

---

[19] Again, this terminological distinction is ultimately of a notational nature. Bauer (2001: 38f), however, sees a 'more fundamental' problem in it, namely that it is "not possible to tell at the point when a word is coined whether it will turn out to be a nonce word or a neologism" – so that "a term is required which is neutral with regard to the diachronic implications that these terms have," and he proposes we use 'coinage' as such a term. I choose to differ here, on two counts: a) 'coinage' to me is not free from diachronic implications either, due to its connotations of intended permanence (a coin, once 'minted', doesn't suddenly drop out of existence again, whereas nonce-formations typically do – see below); and b) I'd say that at the point when a word is formed it *is* a nonce-formation *per definitionem*, the question is only what happens next. If we need a neutral cover term, why not simply speak of 'new formation' for both (with systematic ambiguity): absolutely new (nonce) and relatively new diachronically (neologism). The latter *is*, after all, a fuzzy concept. Admittedly, though, Bauer (2001) only rejects the nonce vs. neologism distinction as one irrelevant to the notions of productivity and morphological structure – which is probably justified. For the notions of lexicalization and institutionalization, however, the distinction is crucial.

occurrence only. 'Nonce' *can* be the first stage in a longer life-span of a word but need not be – and mostly it is also the last stage.

Neologisms are also awkward because they have to be considered a transitional phenomenon – no longer a nonce-formation, but not yet a fully institutionalized member of the lexicon either.

Similar distinctions of perspective have to be observed here to the ones regarding 'established' words in section 3.2 (type-familiar vs. item-familiar, institutionalized, lexicalized). A formation may be institutionalized in the language of one speaker's (subset of a) speech community, but may be perceived by an 'outsider' listener as a nonce-formation, even though it wasn't one from the speaker's point of view – and vice versa, i.e. a speaker may form a nonce-formation (from his/her perspective) but it may already be part of a listener's lexicon.

Such uncertainties may apply mostly to nonce-formations which have only begun the transition to the status of neologism. A large proportion of nonce-formations, however, never even make it this far.

4.2 (Non-)Lexicalizability

Quite frequently the assumption can be encountered that any new word-formation is potentially lexicalizable, e.g. Kastovsky (1993: 6): "The output of word-formation processes consist of lexical items. These are […] potentially listable and are in fact more often than not integrated into the permanent vocabulary of a speech community […]."

However, this common assumption is not really supported by the evidence available. While all words that are in the lexicon are positive evidence of words becoming listed (they must have started out as new words at some point), evidence of words that may have been formed at some point but never were listed is hardly quantifiable. We simply haven't got sufficient records of all the words that never made it into the lexicon. To test Kastovsky's assumption that new words "more often than not" end up in permanent vocabulary, however, we may, as a first approximation, attempt to keep track of 'new words' collected in dictionaries of neologisms by checking how many of these end up in standard dictionaries some years later. Kjellmer (2000: 226) summarizes the outcome of such studies as follows: "It appears that neologisms due to semantic change have in general a much better chance of survival than other neologisms, but that in other cases only half or less than half of them stay on in the language."

Half or less is not 'more often than not' – not even for (also formally) new words that were at least temporarily already institutionalized to a degree as neologisms. As regards wholly new formations that didn't even make it to the neologism status, i.e. proper nonce-formations, these are by their very nature even more elusive. Specific studies of this area of word-formation like Hohenhaus (1996), (2000), (2004) suggest that it is far more typical of such new formations not to become lexicalized, not even temporarily as neologisms. Again, hindsight corroborates this: out of the ca. 600 nonce-formations collected as the empirical basis of Hohenhaus (1996), virtually none have been adopted as current permanent vocabulary items.

It is thus worthwhile and sensible to ask whether there are any generalizations possible about formations that can (and do) enter the lexicon and those which apparently don't, and whether there may be systematic reasons for this, rather than assuming that all words are intended for the lexicon (like e.g. Motsch 1977 holds) and ascribing it to historical accident alone if they don't make it there.

Downing (1977) provided an early challenge to this view, exploring characteristics of new nominal compounds, including ones that have a systematic impact on lexicalizability – in particular: permanence or fortuitousness of the underlying relationship between the constituents. Her well-known example of a compound lacking a suitable lexicalizable underlying relationship is *apple-juice seat* – used to refer to a seat in front of which a glass of apple juice was placed at a particular social event ('refer' in the sense of single out, not designate). This is why Downing (1977: 819) speaks of 'deictic compounds' in such cases. And it seems to be the context-dependency of such forms that make them non-lexicalizable.[20]

This was further explored in Hohenhaus (1996), (1998), (2005b) and a general theory of (non-)lexicalizability proposed. It entails an important conceptual distinction between *potential listeme* (i.e. a possible word purely from the point of view of the lexicon) and *possible word* in the sense of a word that can be formed and used in performance (as an X^0, in generative terminology) irrespective of whether or not it could also be listed. Genericness of meaning vs. individual (local) context-dependency seems to be a crucial factor, but there are also whole productive patterns of formation that only yield non-lexicalizable, though perfectly regular, formations, i.e. ones that have no counterpart in the lexicon at all:

For instance, Bauer (1983: 90) remarks on so-called *expletive infixation* (such as *abso-bloody-lutely*) that "words produced by such infixing never seem to become established." Other such types include what I have dubbed *dummy compounds* and *identical constituent compounds*. The former are a type of highly productive text-deictic compounds employing *thing* or *business* as virtually empty pro-forms in head position – such as *vacation thing*, *greengages business*, used to refer back to stretches of previous co(n)text in which something was said about vacations and greengages, respectively (for contexts and extensive discussion see Hohenhaus

---

[20] Note that in such cases we are talking about particular pairings of form and meaning. So when we speak about the non-lexicalization of a given form, this is really shorthand for saying: non-lexicalization of this particular form with this particular meaning. It is not necessarily to say that the form as such (paired with perhaps quite a different meaning) is in principle non-lexicalizable. We also have to observe the difference between naming and referring here. The fact that normally *apple-juice seat* is non-existent in extra-linguistic reality (see Štekauer 2002: 110) is not so crucial alone. In its specific context such an object *did* 'exist' (temporarily!) – the point is rather that it didn't exist as a generally 'name-worthy CATEGORY' (Downing 1977: 823, emphasis in the original). While the permanent lexicon contains permanent names, it is not the case that words are formed exclusively to fill naming needs. Nonce word-formation in particular can fulfil a range of functions quite different from naming, including purely fortuitously deictic and meta-communicative ones that have nothing to do with permanent categories and are thus much more syntactic in function (cf. also Kastovsky's 1982b distinction between naming and 'syntactic recategorization'). Such functional considerations, however, are beyond the scope of this chapter (but see Hohenhaus 1996, chapter 5).

2000, 1996: 281ff).[21] The latter consist of repeated identical constituents, e.g. *instant-instant* or *pain-pain*, as in the following contexts:

(3)   Felix: Tim! I'd be careful. That's instant glue you're using.
        Tim: It's not like instant-instant. [22]

            Adam: [referring to his partner who is undergoing fertility treatment] … that she's going through all this pain. I mean not just emotional pain, but pain-pain. [23]

Again, these serve a meta-communicative function by regularly prompting a 'prototype-type' interpretation: '(an) XX = (a) proper/real X' (for nouns, for adjectives and verbs: 'XX = properly/absolutely/completely X'). In a corpus-based study (Hohenhaus 2004), it was confirmed that, while the type is productive, its products, the individual compounds themselves, are hapaxes. If none are established, the pattern has to be regarded as being exclusively a nonce type.[24]

What is important to stress here overall is: some new formations, even if productively formed and used in an X^0 position in performance, are not necessarily potential listemes at the same time. What is possible 'for keeps' in the lexicon is not identical to what is possible as a formation.

The concept of non-lexicalizability is obviously at odds with models in which all outputs of word-formation first have to enter the lexicon before being usable in syntax (as in Halle 1973 or Štekauer 2002). In my model (Hohenhaus 1996: 249) word-formation can feed directly into syntax, bypassing the lexicon altogether, which is more in line with e.g. Aronoff & Anshen's (1998: 237) emphasizing that, interdependent as morphology and the lexicon may be, they constitute two different sources of words. A theory of word-formation should reflect this.

4.3 What is in the (mental) lexicon and how does it get there?

As we have seen above, the question of what ends up in the lexicon can be approached by 'negative elimination', as it were, by trying to identify what cannot enter the lexicon. While this is still an advancement over approaches that leave this basically to historical chance alone, it still does not describe how items that *are* potential listemes do become actual listemes.

Herbermann (1981: 325ff) addresses this question more directly, by proposing specific procedures by which a newly created complex lexical item can then be promoted to the status of 'lexeme' via successive stages of introduction, reiteration,

---

[21] Context-dependency is clearly at work here as well: e.g. *vacation thing* 'means' in its context: 'what you told me at a given time/place about (a particular) vacation'. The deictic specificity of reference of different sorts (personal, temporal, spatial) makes this a non-generic form-meaning pairing, and thus a non-lexicalizable one in any case.

[22] From the television series *Home Improvement*, episode 69; a transcript can be found at: <http://www.hiarchive.co.uk/script.php?s=3&e=20>

[23] From the television series *Cold Feet*, third series, episode 3, ITV, aired on 26 November 2000.

[24] See, however, Hohenhaus (2005b) for some caveats regarding the equivalent phenomenon in German.

transposition to other 'texts' etc.; but while Herbermann's considerations may be plausible enough in themselves, they do not go very far in explaining what happens to an individual speaker's mental lexicon. They are rather speculations on how a new formation may *spread* as a new lexeme from the original speaker to a smaller or larger speech community, i.e. they concern institutionalization of neologisms. But if words enter a collective lexicon, they have to enter the individuals' lexicons. How is this achieved? Is there a specific 'operation' that creates new lexical entries? This is sometimes implied (cf. e.g. Toman 1983: 4f, 38f, Lieber 1992: 159). However, psycholinguistically more flexible approaches may be preferable; cf. for instance Meys' (1985: 77) speculations about "mental traces left by productive/interpretative occurrences" which would gradually facilitate listing according to frequency. Especially if we conceive of the mental lexicon as a (neural) network (cf. Aitchison 1994: 228), this seems quite plausible, as higher frequency could be understood as strengthening the links between nodes.[25]

And if there is *gradual listing*, we would of course expect to find degrees of (semi-)listedness. This would also be in line with Jackendoff (1997: 231, note 11): "[There needn't be] a strict cutoff in frequency between stored forms and forms generated on-line. Rather I would expect a cline in accessibility for stored forms, including the usual differential between recognition (better) and recall (worse)."

Assessing psychological reality empirically, however, has always been a problem. Word recognition tests and other psycholinguistic experiments may inform us about whether or not, or to what degree, complex words are stored in individuals' mental lexicons and whether they are stored as whole-form entries or with morphological structure/decomposition playing a role. However, the experimental evidence, as McQueen & Cutler (1998) or Bauer (2001: chapter 4) summarize it, is still somewhat mixed.

Overall, theoretical models that focus on conceptual simplicity and thus allowing only for an impoverished-entry lexicon of minimal signs appear least corroborated by the psycholinguistic findings so far, while *full(er)-entry models*, especially *network* models (cf. also Lipka 2002: 197ff) seem to fare better. Clearly, psycholinguistics still has a long way to go until we get a fuller understanding of the real nature of the mental lexicon, but there is an emerging thread: both simple and a sizeable subset of complex words and word-forms, even including regular ones (e.g. high-frequency past tense forms), are mentally stored. And whether or not a word-formation is stored depends on a complex balance of factors other than (or in addition to) idiosyncrasy, including, apart from meaning and morphological make-up, indeed frequency (see above), but also aspects such as *family size effects* (word recognition tends to be better if the word shares a constituent with many other words in the lexicon; cf. De Jong, Schreuder & Baayen 2000).

A particularly compelling argument for storage rather than rule-governed on-line composition comes from language acquisition, namely from the so-called *critical-mass hypothesis*: "acquirers need a large number of words at their command before they can generalize over them" (Bauer 2001: 114). Thus a word such as *orange juice* is likely to be learned (stored) as a whole before it is even felt to be transparent (see

---

[25] But further factors have to be considered – cf. Hohenhaus (1998), Kjellmer (2000)

Bauer 2001: 122). That is to say lexicalization (in the synchronic sense) precedes the emergence of the means for on-line composition. The question then is whether words once stored should later become 'unlisted' again. Given that word-memory has to continue through life in any case (new acronyms, new loan words, new terminology, new names, etc.), it seems unlikely that 'unlisting' and switching to on-line (de-)composition should automatically be preferable or more economical, although for words learned in later stages the role of morphological rules in lexical access does seem to increase (cf. Bauer 2001: 212).

4.4 Unpredictable & playful formations, analogy, fads, and new developments

As so often, the devil is the details of concrete cases – and just a few shall be mentioned briefly here:

Recall the cases of 'creative' deviant formations such as *oid-y* or *greenth*. It could be assumed that such cases remain individual 'quirks' or one-offs. Consider, however, a case such as German *unkaputtbar* (literally 'un-broken-able'), which is clearly morphologically deviant because *un-* + A + *-bar* is not an available pattern (like for English *-able*, the base for suffixation by *-bar* has to be a transitive verb). Nevertheless, the word sprung into almost instantaneous currency through an advertising campaign in Germany in 1990 by a big American soft-drinks corporation (of brown fizzy liquid fame) showing one of the then newly introduced plastic bottles (of this brand) together with the single word *unkaputtbar*.[26] Apparently this was suffcent to institutionalize the formation enough to become usable outside its context.[27] On the other hand, in widespread use as it may be, it is not fully accepted – in fact it is a regular target for purist criticism,[28] in particular on the grounds that a perfectly regular word for the concept already existed: *unzerbrechlich*. Thus even though its validity is contested, the deviant formation has to be taken to be established, due to, at least initially, media power rather than normal spread through a speech community. Rather it was 'artificially' institutionalized rapidly and despite lacking acceptance.

Another problematic case for the conception of normal institutionalization and lexicalization as gradual enrichment of vocabulary are sudden bursts of morphological 'fads'. One such case is/was the craze in Germany around the year 2000 of forming more and more complex words roughly in analogy to the initial model of *Warmduscher* (literally 'warm-shower-er') such as *Auf-dem-Schrank-Staubwischer* ('on-top-of-the-wardrobe-duster') or *Bei-Mami-Wäscher* ('laundry-at-Mom's-doer'). They are all intended to mean ultimately the same (i.e. naming needs play hardly any role here): something between 'wimp', 'pathetically conventional or

---

[26] Cf. <http: //www.slogans.de/>; Apparently there was also a television clip in which such a bottle was shown tumbling down a staircase without suffering damage.

[27] An Internet search engine returned over 14,000 hits (in July 2004); the word is now also used in all manner of contexts in the sense of 'unbreakable' – including the title of a recent CD by a German pop singer!

[28] Cf. for instance <http: //uleuschner.bei.t-online.de/rezensionen/ri9610szymanski.htm>, or the relevant discussion thread in the chat forum of the purist association VDS: <http: //www.vds-ev.de/forum/>.

pedantic person' or 'do-gooder'. Some specimens clearly overstretch the limits of wellformedness (e.g. *Hochzeitstagdrandenker* 'anniversary-remember-it-er' – containing the object of the verb simultaneously in lexical and pronominal form internal to the compound!). However, many will have been institutionalized for a sizeable speech-community (much larger than the small-group intimate varieties referred to in section 3.2 as more likely domains for such phenomena) – there were clubs and popular dedicated websites indulging in this particular exercise in word-play. Ultimately, few speakers will have escaped contact with the phenomenon. A few years on, though, the fad has largely ebbed away and hardly any products of it survive (perhaps only the original *Warmduscher*), i.e. the 'vocabulary' has not been enriched in any permanent sense at all. De-institutionalization followed institutionalization rapidly.[29]

Finally, consider all those new phenomena in computer-mediated language (cf. Crystal 2001, Hohenhaus 2005a), such as *reduplications* like *nodnod* or *wavewave* used as so-called *emotes* in certain genres. These seem to be formed productively, but at least a certain subset will have conventionalized to the degree of full institutionalization. In chat and text messaging (SMS), where brevity is imperative, acronyms and other methods of shortening are thriving, giving rise to a mix of conventionalized forms such as *lol* (for 'laugh out loud') or *CU l8r* (for 'see you later') and generally applicable technique (non-lexicalized, rule-governed) – e.g. the replacement of any homophone syllable by numbers (esp. *1*, *2*, *4*, *8* and *9*). Furthermore there is the use of iconic signs, the best known example being the *emoticon* : –) but again there is a fuzzy cline of institutionalization: ;–) for 'wink' and : –(for 'unhappy' are probably known to most e-language users, but rarer constructions such as :–@ for 'screaming' are probably only familiar to comparatively fewer, seasoned 'netties', where they may even become in-group identifiers. These are still volatile areas of language where standardization may just be setting in and which need to be observed more to see what influence they may exert on general word-formation and the lexicon.

4.5 Lexicalization beyond words

So far we have only considered lexicalization and institutionalization as applying to words – as is appropriate in a handbook of *word-formation*. However, a few brief remarks are in place about the question whether there are units other than words in the lexicon as well. And indeed there are good arguments for such an assumption.

Early approaches, notably Weinreich (1969), proposed the integration of a separate list of idioms *alongside* the (word-)lexicon proper. Furthermore, the fact that larger-than-word objects can also play a role in word-formation, in particular in phrasal compounds, lead to various approaches of integrating this, either wholesale

---

[29] Admittedly, this is not easy to prove empirically, as the products remain accessible in written records of the fad. Also, quite a few items in such lists (e.g. <http: //www.ablachen.de/weichei.htm>) are actually normal words, such as *Fahrschüler* 'learner driver', that have long been established in their primary, neutral meaning, but were given a connotational twist simply by inclusion in such a list alongside genuinely new formations.

(e.g. Lieber 1988, 1992, Bauer 1983) or by isolating the scope of this apparently syntactic element within word-formation proper (cf. Hohenhaus 1996: 218-228).[30]

A particularly interesting proposal is the integrational approach of Jackendoff (1997). The empirical basis for his endeavour comes from his '*Wheel of Fortune* Corpus' – a collection (compiled over a few months) of ca. 600 fixed expressions used in the game show of the same name, where contestants have to solve puzzles as quickly as possible from the step-by-step disclosure of letters. What is striking is that there seems to be little strain (on the part of the programme makers) in coming up with and (on the part of the contestants) recognizing vast amounts of such strings (and the show has been on air for many years six days a week). Strings in question include not only idioms proper such as *to throw in the towel* or *to eat humble pie*, but also *clichés* such as *we're doing everything humanely possible* or *gimme a break*, and well-known quotations such as *may the Force be with you* or *beam me up, Scotty*. Quite obviously, then, "there are too many idioms and other fixed expressions for us to simply disregard them as phenomena on the margin of grammar" (Jackendoff 1997: 177), and it is even possible that "their number is of the same order of magnitude as the single words of the vocabulary" (Jackendoff 1997: 156). The reason that such fixed expressions have typically been neglected esp. in generative theories (as 'not core grammar') probably does stem, as Jackendoff (1997: 153, 157) surmises, from the simple conception of the lexicon as the source of X^os at the point of lexical insertion only. Having replaced that notion (see section 2.3) with a model of tripartite parallel lexical licensing, Jackendoff (1997) manages to integrate listemes of all formats neatly.

Suffice it to say here, by way of a conclusion: it is clear that the lexicon has to be more than a simple list of 'words'. Rote-learning, memorized building blocks of various sizes and associations between them, form a large and integral part of 'lexical knowledge' alongside (competence-)knowledge of morphemes plus the productive morphological rule-system for on-line (de-)composition of complex words. In short: lexicalization/listing is of great relevance even beyond word-formation!

Peter Hohenhaus
University of Nottingham
Department of German
Trent Building
University Park
University of Nottingham
Nottingham NG7 2RD, UK
e-mail: Peter.Hohenhaus@nottingham.ac.uk

---

[30] More recent evidence of the importance of larger-than-word memorized building-blocks of language also comes from Wray (2002).

REFERENCES

Aitchison, Jean. 1994. *Words in the Mind – an introduction to the mental lexicon.* 2nd edition, Oxford: Blackwell.

Aronoff, Mark. 1976. *Word Formation in Generative Grammar.* Cambridge, Mass.: MIT Press.

Aronoff, Mark. 1988. Review of: Di Sciullo/Williams (1987). *Language* 64, 766-770.

Aronoff, Mark and Anshen, Frank. 1998. "Morphology and the Lexicon." In: A. Spencer and A. Zwicky (eds.), 237-247.

Bauer, Laurie. 1983. *English Word-formation.* Cambridge: Cambridge University Press.

Bauer, Laurie. 2001. *Morphological Productivity.* Cambridge: Cambridge University Press.

Brekle, Herbert and Kastovsky, Dieter (eds.). 1977. *Perspektiven der Wortbildungsforschung. Beiträge zum Wuppertaler Wortbildungskolloquium vom 9. – 10 Juli 1976. Anläßlich des 70. Geburtstages von Hans Marchand am 1. Oktober 1977.* Bonn: Bouvier.

Coseriu, Eugenio. 1967. "Lexikalische Solidaritäten." *Poetica* 1, 293-203.

Crystal, David. 2001. *Language and the Internet.* Cambridge: Cambridge University Press.

De Jong, Nivja H.; Schreuder, Robert; and Baayen, Harald, R. 2000. "The morphological family size effect and morphology." *Language and Cognitive Processes* 15, 329-365

Di Sciullo, Anna-Maria and Williams, Edwin. 1987. *On the Definition of Word.* Cambridge, Mass.: MIT Press.

Downing, Pamela. 1977. "On the creation and use of English compound nouns." *Language* 53, 810-842.

Halle, Morris. 1973. "Prolegomena to a theory of word-formation." *Linguistic Inquiry* 4, 3-16.

Herbermann, Clemens-Peter. 1981. *Wort, Basis, Lexem und die Grenze zwischen Lexikon und Grammatik.* München: Fink.

Heringer, Hans Jürgen. 1984. "Wortbildung: Sinn aus dem Chaos." *Deutsche Sprache* 12, 1-13.

Hohenhaus, Peter. 1996. *Ad-hoc Wortbildung – Terminologie, Typologie und Theorie kreativer Wortbildung im Englischen.* Frankfurt/M., Berlin, Bern, New York, Paris, Wien: Peter Lang.

Hohenhaus, Peter. 1998. "Non-Lexicalizability as a Characteristic Feature of Nonce Word-Formation in English and German." *Lexicology* 4/2, 237-280.

Hohenhaus, Peter. 2000. "An Overlooked Type of Word-Formation: Dummy-compounds in German and English." In: Ch. Hall and D. Rock (eds.), *German Studies towards the Millennium.* Oxford, Bern, Berlin, Bruxelles, Frankfurt/M., New York, Wien, 241-260.

Hohenhaus, Peter. 2004. "Identical Constituent Compounds – a Corpus-based Study." *Folia Linguistica* 38/3-4, 297-331.

Hohenhaus, Peter. 2005a. "Elements of traditional and 'reverse' purism in relation to computer-mediated communication." In: N. Langer and W. Davies (eds.), *Linguistic Purism in the Germanic Languages.* Berlin, New York: De Gruyter, 186-202.

Hohenhaus, Peter. 2005b. "Where Word-Formation cannot Extend the Vocabulary - Creativity, Productivity and the Lexicon in Synchronic and Diachronic Morphology." In: D. Kastovsky and A. Mettinger (eds.), *Lexical Change and the Genesis of the English Vocabulary.* Berlin, New York: Mouton de Gruyter.

Hughes, Geoffrey. 2000. *A History of English Words.* Oxford: Blackwell.

Jackendoff, Ray. 1975. "Morphological and Semantic Regularities in the Lexicon." *Language* 51, 639-671.

Jackendoff, Ray. 1997. *The Architecture of the Language Faculty.* Cambridge, Mass.: MIT Press.

Kastovsky, Dieter. 1982a. *Semantik und Wortbildung.* Düsseldorf, Bern, München: Bagel/Francke.

Kastovsky, Dieter. 1982b. "Word-Formation: A Functional View." *Folia Linguistica* 16, 181-198.

Kastovsky, Dieter. 1993. *The Place of WF in Grammar: A Historical Survey.* ICOLS.

Kjellmer, Göran. 2000. "Potential Words." *Word* 51/2, 205-228.

Leech, Geoffrey. 1974. *Semantics.* Harmondsworth: Penguin.

Lehmann, Christian. 2002. *New reflections on grammaticalization and lexicalization.* In: I. Wischer and G. Diewald (eds.), *New Reflections on Grammaticalization.* Amsterdam, Philadelphia: John Benjamins, 1-18.

Lieber, Rochelle. 1981. *On the Organization of the Lexicon.* Bloomington, Indiana: Indiana University Linguistics Club.

Lieber, Rochelle. 1988. "Phrasal Compounds in English and the Morphology-Syntax Interface." In: *Papers from the 24th Annual Regional Meeting of the Chicago Linguistics Society. Part Two: Parasession on Agreement in Grammatical Theory* (= *CLS* 14), 175-184.

Lieber, Rochelle. 1992. *Deconstructing Morphology*. Chicago, London: University of Chicago Press.

Lipka, Leonhard. 1977. "Lexikalisierung, Idiomatisierung und Hypostatisierung als Probleme einer synchronen Wortbildungslehre." In: H. Brekle and D. Kastovsky (eds.),1977, 155-163.

Lipka, Leonhard. 1981. "Zur Lexikalisierung im Deutschen und Englischen." In: L. Lipka and H. Günther (eds.), *Wortbildung*. Darmstadt: Wissenschaftliche Buchgesellschaft, 119-132.

Lipka, Leonhard. 2002. *English Lexicology: Lexical structure, word semantics & word-formation*. Tübingen: Narr.

Lyons, John. 1977. *Semantics*. Cambridge: Cambridge University Press.

McQueen, James M. and Cutler, Anne. 1998. "Morphology in Word Recognition." In: A. Spencer and A. Zwicky (eds.),1998, 406-427.

McWhorter, John. 2001. *The Power of Babel – a Natural History of Language*. New York: Henry Holt/Times.

Meys, Willem, 1975. *Compound Adjectives in English and the Ideal Speaker-Listener*. Amsterdam: North Holland.

Meys, Willem. 1985. "Morphological meaning and the structure of the mental lexicon." In: A. J. Geer, P. Hoppenbrouwers, A. M. Seuren, and T. Weijters (eds.), *Meaning and the Lexicon*. Dordrecht: Foris, 76–85.

Motsch, Wolfgang. 1977. "Ein Plädoyer für die Beschreibung von Wortbildungen auf der Grundlage des Lexikons." In: H. Brekle and D. Kastovsky (eds.), 180-202.

Pinker, Steven. 1994. *The Language Instinct*. London: Penguin.

Selkirk, Elizabeth. 1982. *The Syntax of Words*. Cambridge, Mass.: MIT Press.

Spencer, Andrew and Zwicky, Arnold. 1998. *The Handbook of Morphology*. Oxford: Blackwell.

Štekauer, Pavol. 2000. *English Word-Formation – A History of Research (1960-1995)*. Tübingen: Gunter Narr.

Štekauer, Pavol. 2002. "On the Theory of Neologisms and Nonce-formations." *Australian Journal of Linguistics* 22/1, 97-112.

Toman, Jindřich, 1983. *Wortsyntax*. Tübingen: Niemeyer.

Weinreich, Uriel. 1969. "Problems in the Analysis of Idioms." In: J. Puhvel (ed.), *Substance and Structure of Language*. Los Angeles, Berkeley: University of California Press.

Welte, Werner. 1996. *Englische Morphologie und Wortbildung*. 2nd edition. Frankfurt/M, Berlin, Bern, New York, Paris, Wien: Peter Lang.

Wray, Alison. 2002. *Formulaic Language and the Lexicon*. Cambridge: Cambridge University Press.