Back-formation in a new theoretical universe

Livio Gaeta & Fabio Montermini
University of Turin & CNRS / University of Toulouse Jean Jaurès

Traditionally, the phenomenon of so-called back-formation occupies only some lines in introductory textbooks in word-formation, and, with few exceptions (Becker 1993, Plag 2003: 27, Rainer 2004), it is discussed even more sparsely in theoretical works, often as cases of "reanalysis/reinterpretation involv[ing] some analogical pressures, especially when the reanalysis is induced by models that exist elsewhere in the language" (Joseph 2001, cf. also Mel'čuk 2001: 532). It is not rare, moreover, that back-formation is included among minor word-formation processes on a par with extragrammatical or marginal morphological phenomena like blending, clipping and the like (Bauer 1983: 232, Lieber 2005: 375, see Štekauer 2015 for a survey).

A further cue of the marginality of the phenomenon is the fact that its borders are fuzzy, and that it is not clear what we should consider genuine cases of back-formation, by contrast with 'canonical' derivation. Classic examples include most often cases of affix (suffix) deletion as *burglar → burgle* or *baby-sitter → baby-sit*. However, both a purely formal criterion or an etymological / diachronic one (which word is attested first) are probably insufficient to provide a full account of the phenomenon. It is likely that semantic and structural criteria also play a role in the interpretation of a morphological relation as being an instance of back-formation.

On the other hand, recent research trends in morphology in various theoretical frameworks have shifted the focus from purely derivational rules to lexical / derivational networks or paradigms. As a consequence, the very role of directionality in word-formation (and more generally in linguistics) has been challenged. In particular, multidirectionality and multiple motivation have been identified as constitutive properties for many derived lexemes. Concurrently, analogy has progressively been recognized as a driving force for derivation, losing its status of a marginal, unpredictable, phenomenon. In this picture, one can wonder whether "back-formation" can still be considered a theoretical relevant concept, or rather a merely descriptive label.

We call for proposals devoted to both theoretical issues and concrete case studies of back-formation in any language, and theoretical perspective. A non-exhaustive list of possible issues to be addressed is the following:

- What is back-formation? Does it correspond to a theoretically relevant notion for linguistics (morphology, lexical semantics, lexicography…)?
- Is back-formation a well-defined set of phenomena? How to determine its borders and content?
- What are the properties of back-formation? What is its relation to subtraction, clipping and other similar phenomena?
- Should back-formation as a diachronic phenomenon be distinguished from back-derivation as a synchronic process on a par with other word-formation processes?
- Is back-formation a universal phenomenon or is it limited to a subset of languages (e.g. to agglutinating morphology)?
- Is back-formation limited to derivation, or should it include inflectional analogical phenomena (e.g. French *château*$_{SG}$ 'castle' from etymological *châteaux*$_{PL}$ 'castles'; Colloquial Italian *perplimere* 'to perplex' from the pseudo-past participle *perplesso* 'perplexed')?

- Are there cognitive / acquisitional cues that allow distinguishing back-formation from other morphological phenomena?
- What can corpus linguistics – and more generally electronically available data-bases – tell us with regard to the consistence and to the measurability (productivity, frequency, etc.) of back-formation?

Scientific Committee

Mark Aronoff
Laurie Bauer
Wolfgang U. Dressler
Maria Grossmann
Antonio Fábregas
Bernard Fradin
Brian Joseph
Rochelle Lieber
Elisa Mattiello
Fiammetta Namer
Susan Olsen
Franz Rainer
Pavel Štekauer
Anna M. Thornton
Petra M. Vogel

References

Bauer, Laurie (1983), *English word-formation*. Cambridge: Cambridge University Press.

Becker, Thomas (1993), Back-formation, cross-formation, and "bracketing paradoxes" in paradigmatic morphology. *Yearbook of Morphology 1993*, 1-25.

Joseph, Brian D. (2001), Diachronic Morphology. In Andrew Spencer & Arnold Zwicky (eds.), *The Handbook of Morphology*. Oxford: Blackwell.

Lieber, Rochelle (2005), English word-formation processes. In Pavol Štekauer & Rochelle Lieber (eds.), *Handbook of Word Formation*. Dordrecht: Springer, 375-427.

Mel'čuk, Igor (2001), Formal processes. In Geert Booij, Christian Lehmann, Joachim Mugdan (eds.), *Morphologie / Morphology*. Berlin / New York: de Gruyter, vol 1, 523-535.

Plag, Ingo (2003), *English word-formation*. Cambridge: Cambridge University Press.

Rainer, Franz (2004), Retroformazione. In Maria Grossmann & Franz Rainer (eds.), *La formazione delle parole in italiano*. Tübingen: Niemeyer, 493-497.

Štekauer, Pavol (2015), Backformation. In Peter O. Müller, Ingeborg Ohnheiser, Susan Olsen & Franz Rainer (eds.), *Word-Formation. An International Handbook of the Languages of Europe*. Berlin / New York: de Gruyter, vol 1, 340-352.

Forward-formation and back-formation in a Relational Morphology perspective

Martin Haspelmath
Max Planck Institute for Evolutionary Anthropology
haspelmath@gmail.com

Giving special attention to back-formation makes good sense in a context where forward-formation is regarded as the default, and in the informal practice, this is surely the case. We often talk about morphological patterns as consisting of the "addition" of an element to a base, as the "combination" of several elements, and so on. Process metaphors permeate our ways of talking, and sometimes also our ways of thinking about grammatical patterns. But it has long been known that directional processes can be problematic (Hyockett 1954), and nondirectional alternatives have been advanced (e.g. Becker 1993, Bochner 1993, Plag 2003). Recently, a fully explicit proposal for a nondirectional model has been made under the heading of "Relational Morphology" (RM; Jackendoff & Audring 2020).

Two further traditional ideas are constitutive for the special status of back-formation: (i) that derivation is a different kind of morphology, and (ii) that productivity is the default so that its absence needs to be explained. In the RM perspective, these ideas do not play an important role anymore: On the one hand, inflection and derivation (and even syntax) are treated in much the same way, in terms of schemas of different degrees of generality, with interface links among the components of the Parallel Architecture and relational links between related schemas. On the other hand, limited productivity is a separate matter in RM, not treated as directly derivable from other aspects of the model (e.g. from an architectural difference between morphosyntax for productive patterns, and the lexicon for unproductive patterns, as in other models).

The RM perspective thus does not treat back-formation as special, and its seeming rarity must explained in different terms or treated as accidental. And in fact "inflectional back-formation" (or "syntactic back-formation") does not seem to be unusual at all and is typically not even noticed – it appears that the unusualness of some types of back-formation mostly boils down to limited productivity. Štekauer (2015) points to the tension with the principle of diagrammatic iconicity, and similarly Anderson (2020) highlights the unusualness of "semantically subtractive morphology" (which may result in additive back-formation) – but these semantic effects may be best treated from the perspective of frequency-induced efficient coding (Haspelmath 2021).

References

Anderson, Stephen. 2020. Semantically subtractive morphology. In Körtvélyessy, Lívia & Štekauer, Pavol (eds.), *Complex words: Advances in morphology*, 36–54. Cambridge: Cambridge University Press.

Becker, Thomas. 1993. Back-formation, cross-formation, and 'bracketing paradoxes' in paradigmatic morphology. In Booij, Geert & van Marle, Jaap (eds.), *Yearbook of Morphology 1993*, 1–25. Dordrecht: Springer. (DOI: 10.1007/978-94-017-3712-8_1)

Bochner, Harry. 1993. *Simplicity in generative morphology*. Berlin: Mouton de Gruyter.

Haspelmath, Martin. 2021. Explaining grammatical coding asymmetries: Form-frequency correspondences and predictability. *Journal of Linguistics* 57(3). 605–633. (doi:10.1017/S0022226720000535)

Hockett, Charles F. 1954. Two models of grammatical description. *Word* 10(2–3). 210–234.

Jackendoff, Ray & Audring, Jenny. 2020. *The texture of the lexicon: Relational Morphology and the Parallel Architecture*. Oxford: Oxford University Press.

Plag, Ingo. 2003. *Word-formation in English*. Cambridge: Cambridge University Press.

Štekauer, Pavol. 2015. Backformation. In Müller, Peter O. & Ohnheiser, Ingeborg & Olsen, Susan & Rainer, Franz (eds.), *Word-Formation*, 340–352. Berlin: De Gruyter Mouton. (doi:10.1515/9783110246254-016)

Back-formation as borrowed productive morphology in Basque

Julen Manterola
University of the Basque Country (UPV/EHU)
julen.manterola@ehu.eus

José Ignacio Hualde
University of Illinois Urbana-Champaign
jihualde@gmail.com

The process that leads to the borrowing of suffixes is both frequent and generally well understood. Conversely, the borrowing of back-formation processes is less well documented. This paper describes and analyzes a specific process of creation of nouns from verbs by back-formation in Basque that may have been affected by similar processes present in Romance languages, in particular, Gascon and Spanish.

In Basque we find a rather large group of nouns/adjectives paired with related verbs of the type *luze* 'long' / *luza-tu* 'to lengthen', *oso* 'whole' / *osa-tu* 'to complete'. These pairs are best explained through regular *-tu* derivation. This is a productive morphological pattern, subject to specific vowel alternations: in two-syllable words, the stem-final vowels *-o*, *-e*, *-a* are neutralized in *-a*; on the other hand, high vowels such as *-u* and *-i* do not enter into this alternation (Mitxelena 1961-1977). This phenomenon is not exclusive to N → V derivation, as it is well documented with other derivational suffixes as well as in composition; e.g. *luze* 'long' + *-garri* [SUFF] → *luzagarri* 'prolongation', *baso* 'forest' + *katu* 'cat' → *basa-katu* 'wild cat', *etxe* 'house' + *jaun* 'master' → *etxajaun* 'master of the house', etc. The vowel alternation most likely has its origin in earlier neutralization in /ə/ in this context.

In many other pairs, however, we arguably have the opposite derivation, V → N by back-formation. This is most obvious in the case of numerous borrowings: in *deitu* 'to call, to be named' / *dei* 'call' we must be dealing with back-formation, since in Old Navarrese-Aragonese we only find *deito* 'called, named' (ultimately from Lat. *dictum*), while no corresponding noun **dei* is recorded (EHHE, s.v); similarly, in *asmo ~ asmu* 'thought, intention' / *asma-tu* 'to guess, imagine', only an Old Spanish verb *asmar* exists (from Lat. *ad-aestimare*), and a Romance noun **asmo* is nowhere to be documented. One further example of this process is *uztatu* 'to harvest' → *uzta* 'harvest', where a verb formally related to Gascon *oustà* (< Lat. *augustare*) was most probably borrowed (EHHE, s.v).

Many of these examples witness that the aforementioned vowel alternation is also reverted in the verb-to-noun back-formation process (see *asmatu* → *asmo*). The pattern is, in fact, frequent and productive enough that it has been used by lexicographers to create new nouns and adjectives from verbs; e.g. *askatu* 'to release' → *aske* 'free' (Urgell 2003).

If V → N back formation simply involved the reversal of the phonological changes in the N → V process, we would not expect to have any deverbal nouns in *-u*, since, as mentioned above, nouns and adjectives ending in high vowels do not undergo final-vowel

lowering. But in fact, we do find quite a few. That is, the converse of N/Adj {-a, -e, -o} → V {-a} is actually V {-a} → N/Adj {-a, -e-, -o, -u}. This paper argues that contact with Romance languages may have contributed to the extension of back-formation patterns in Basque, enhancing the intrinsic productivity of deverbal processes in this language and extending morphophonological alternations to new contexts. It should be noticed that in the contact Romance languages, V → N zero-derivation (without the additions of derivational suffixes) is common and productive (see Prat Sabater 2009), e.g. Sp *gast-a-r* 'to spend' → *gast-o* 'expense', *cost-a-r* 'to cost' → *cost-o*, *cost-e* 'cost', Gasc *cost-a-r* → *cost*, potentially allowing for the borrowing or strengthening in Basque of this morphological procedure.

The irregularity under study may be characterized as this way: in some instances where a N has been derived within Basque from a borrowed V, the N ends in *-u*; e.g. Bq *izkriba-tu* 'to write' → *izkribu* 'piece of writing' (where there is a Romance source for the Bq V, but not for the N; cf. Sp *escribir* 'to write', but *un escrito* 'a piece of writing', **un escribo*). The explanation for the final *-u* in deverbal nouns may be the following: Sp nouns in *-o* often have Bq cognates in *-u*, because they were borrowed before the lowering of Lat *-u* in Ibero-Romance and the deletion of this vowel in Gallo-Romane (e.g. Sp *pino*, Gasc *pin*, Bq *pinu* 'pine tree'). In addition, in a number of cases both a noun and its derived verb were borrowed, so that the two languages in contact share both cognates; e.g. Sp *baño* 'bath', *bañar* 'bathe'; Bq *bainu*, *baina-tu*. This created a pattern V *-a(-tu)* : N *-u* that was then sometimes extended to back-formed native Ns; e.g. *aipa-tu* 'to mention' → *aipu* 'a mention'.

Although the borrowing of back-formation processes appears to be uncommon, in the case at hand, the confluence of a number of factors may have facilitated it in Basque, to the point of extending the context of a morphophonological alternation.

References

EHHE = Lakarra, Joseba, Julen Manterola & Iñaki Segurola. 2019. *Euskal Hiztegi Historiko-Etimologikoa* (= *The Basque Historical-Etymological Dictionary*). Bilbo: Euskaltzaindia.

Mitxelena, Luis. 1961-1977. *Fonética Histórica Vasca*. Donostia: Gipuzkoa Provincial Council.

Prat Sabater, Marta. 2009. Estudio diacrónico sobre la derivación regresiva en español. In Joan Rafel (ed.), *Diachronic linguistics*, 239-274. Girona: Documenta Universitaria.

Urgell, B. 2003. Eratorpena Larramendiren hiztegian (I): Eratorbideen analisia [= Derivation in Larramendi's dictionary (I): An analysis of derivational processes]. *ASJU* 37(1), 203-260.

A paradigmatic analysis of backformation in ParaDis

Nabil Hathout [a]   Fiammetta Namer [b]
[a] CLLE, CNRS & Université de Toulouse Jean Jaurès
[b] ATILF, Université de Lorraine & CNRS

From a descriptive point of view, a back-formed word is an "analogically derived by deleting a suffix (or supposed suffix)" (Plag 2003:54). The process connects a complex word with a formally simpler word (see also Becker 1993). Plag's analysis of this is based on bidirectional patterns he calls paradigmatic (Plag 2003:200). Other analyses are summarized in Stekauer (2015). Regardless of the theoretical tradition in which they are situated, these analyses are based on binary processes that involve two (patterns of) lexemes.

In this talk, we show how backformation can be analyzed in a paradigmatic framework of derivation, the ParaDis model (Hathout and Namer 2022). The starting point is the canonical situation illustrated by the (paradigmatic) family of *to build* presented in the first row of the table. This same paradigm contains another family obtained by prefixing the lexemes of the family of *to build* by *re-*.

Verb	N_action	N_agent	A_possib
to build	building	builder	buildable
to rebuild	rebuilding	rebuilder	rebuildable

These two families are perfectly aligned: their lexemes are in the same form relations and the same meaning relations. They belong to the same morphological paradigm and are in correspondence with formal and semantic families that belong to the same formal and semantic paradigms. Let us now consider the *bodybuilding* family.

Verb	N_action	N_agent	A_possib
to bodybuild	bodybuilding	bodybuilder	bodybuildable
to canyoneer	canyoning	canyoneer	canyonable

On the formal level, the relations between its four lexemes are identical to those of the *to build* family. The formal families corresponding to those of *to build* , *to rebuild* and *bodybuilding* belong to the same formal paradigm. On the other hand, the semantic relations in the *bodybuilding* family are similar to those in the *canyoning* family but different from those in the *to build* family because *to bobybuild* means 'to practice / take part in bodybuilding' whereas the relation between *to build (X)* and *building (of X)* could be glossed as 'to perform the building of X'. The consequence of this difference is that the morphological families of *to build* and *bodybuilding* belong to different morphological paradigms because they correspond to different semantic paradigms.

Backformation consists in superposing the morphological paradigms of *to build* and *bodybuilding* in the same derivational paradigm, thus creating a perceptible discrepancy between form and meaning. Morphological paradigm superposition is generally meaning-driven: we superpose morphological paradigms in correspondence with the same semantic paradigm while ignoring formal variations, typically derivational exponents. With backformation, it is the semantic variations that are ignored.

The talk presents a detailed analysis of the semantic and lexical differences that exist between the family of *to build* and that of *bodybuilding*, and focus on the particular role each of these

lexemes plays in its family. We also show that it is possible to analyze in the same way various backformations in French (**thermosouder** 'to heat seal', *thermosoudure* 'heat sealing', *thermosoudeur* 'heat sealer', *thermosoudable* 'heat sealable'), (cf. Namer, 2012) including those that are not the (apparent) result of composition such as (**intuiter** 'to intuit', *intuition* 'intuition', *intuitif* 'intuitive'), (**imprévoir** 'to be short-sighted', *imprévoyant* 'short-sighted', *imprévoyance* 'short-sightedness') or (**hégémoner** 'hegemonize', *hégémonie* 'hegemony', *hégémonique* 'hegemonic') which illustrate the diversity of families containing backformed lexemes. We further discuss the conditions that might explain the appearance of backformed lexemes in families and paradigms.

Bibliography

Becker, T. (1993). Back-formation, cross-formation, and 'bracketing paradoxes' in paradigmatic morphology. *Yearbook of Morphology* **1992**: 1-27.

Hathout, N. and F. Namer (2022). ParaDis: a family and paradigm model. *Morphology* **32**(2):153-195.

Namer, F. (2012). Nominalisation et composition en français: d'où viennent les verbes composés ? *Lexique* **20**: 173-205.

Plag, I. (2003). *Word-formation in English.* Cambridge, Cambridge University Press.

Štekauer, Pavol (2015), Backformation. In Peter O. Müller, Ingeborg Ohnheiser, Susan Olsen &Franz Rainer (eds.), *Word-Formation. An International Handbook of the Languages of Europe.* Berlin / New York: de Gruyter, vol 1, 340-352.

Backformation: looking back on two centuries of research

Franz Rainer

WU Vienna

franz.rainer@wu.ac.at

The phenomenon of backformation was first hinted at in passing in 1827 in Buttmann's grammar of Ancient Greek. That is to say that it has been around since the beginning of linguistics as an academic discipline, and by 1842 it had already received a name, viz. G. *Rückbildung*.

In my introductory talk of the workshop, I will first sketch the terminological side of the matter. The unnecessary multiplication of terminological entities is well-known to be a favourite sport in our discipline, and backformation does not belie this statement. The German leader word was translated into Modern Latin as *derivatio retrograda* in 1868, which then sparked off a series of loan translations in European languages, including German. The French opted for *dérivation en sens inverse* and *dérivation regressive*, both of which also had their followers. All in all, I have counted over a dozen of terms, whose mutual etymological relationships will be made explicit.

Much more interesting, of course, is the conceptual side of the matter. Right from the beginning, controversies arose on the precise nature of the phenomenon and its delimitation. A broad consensus soon emerged that backformation was just another manifestation of analogy, but the explication of the notion of analogy involved has been more controversial: is it proportional analogy, abduction, or something else? Or does back-formation consist in the reversal of a (bidirectional) rule of word formation? Does backformation necessarily involve subtraction? If not, how are we to distinguish backformation and conversion? A particularly relevant issue since so-called "postverbal nouns" (L. *pugna*), which you commonly find in the chapter on conversion in modern treatises, were among the first phenomena to be discussed under the heading of backformation. Or is backformation just a subspecies of conversion/zero-derivation? Does backformation necessarily involve reanalysis, or does it reduce to a subspecies thereof? Are all backformations of diachronic relevance only? How can backformations be identified? Is backformation in derivation distinct from backformation in inflection?

Questions over questions that, in the second part of my talk, I will treat from a historiographic angle, passing from argument to counter-argument to counter-counter-argument... I hope that such a stroll through the history of research will refresh our memories and thereby allow distinguishing better true novelties from heated-up goulash.

Evidence for backformation in German and Slovak

Pius ten Hacken & Renáta Panocová
Leopold-Franzens-Universität Innsbruck & Pavol Jozef Šafárik University in Košice
pius.ten-hacken@uibk.ac.at, renata.panocova@upjs.sk

Backformation is the formation of a less complex word on the basis of a more complex one. An example is the formation of the English verb *compensate* from the noun *compensation*. A central problem in determining backformation in such cases is that the resulting pair is the same as when the noun is formed from the verb. For English, OED (2000-2021) gives extensive information about first attestation dates of words, which can be used to identify individual cases and general trends, as demonstrated by ten Hacken & Panocová (2022). Here we address the situation in German and in Slovak, two languages for which no such information is available.

In German, the suffix *-ation* occurs only with verbs in *-ieren*. It competes with the suffix *-ung*, which can apply to a larger range of verbs. This gives rise to triples such as *Organisation - organisieren - Organisierung*. There are three plausible scenarios for the origin of such a triple. In the first, *Organisation* is borrowed and reanalysed as a complex word, making *organisieren* a backformation, which may then be the base for *Organisierung*. In the second scenario, both *Organisation* and *organisieren* are borrowed. This gives rise to the reanalysis of *Organisation* as a complex word. There is no backformation, but *Organisierung* can be formed on the basis of the verb. In the third scenario, only *organisieren* is borrowed. Both nouns are formed from the verb. This cannot be the only scenario, because the origin of the suffix *-ation* remains unexplained.

In order to gather evidence for the different scenarios, we collected nouns in *-ation* from DeReWo (2012). We found 706 nouns with a corresponding verb in *-ieren*. For these pairs, we also checked whether there are nouns in *-ung*. For all these nouns and verbs, we looked up the frequency in the DeReKo corpus and considered how these frequencies support each of the scenarios described above.

In Slovak, the corresponding suffix is *-ácia*. This suffix is in competition with the native suffix *-nie* which can be used with a large number of verbs. The nouns in *-ácia* occur with verbs in *-ovať*. We collected these nouns from SNK (2020) including their frequencies. There are 459 nouns in *-ácia* for which corresponding nouns in *-nie* and verbs in *-ovať* are attested as well. We found that nouns in *-ácia* tend to have higher frequencies than their corresponding verbs (70% of cases). This effect is even stronger for nouns in *-nie*, where 82 % nouns in *-ácia* have higher frequencies.

On the basis of the analysis of German and Slovak data, we develop an argument to determine which of the three scenarios is most likely for each case. The frequency data from Slovak tend to point to the first scenario, making backformation a relevant factor in the origin of the verbs. In German the situation is more complex, because nouns in *-ung* have a different relationship to the verb. However, we argue for a significant role of backformation also in German.

References

DeReWo (2012), derewo-v-ww-bll-320000g 2012-12-31-1.0, Mannheim: Institut für Deutsche Sprache.

DeReKo (2005-2021), *Deutsches Referenzkorpus*, Mannheim: Institut für Deutsche Sprache, www.ids-mannheim.de/DeReKo

ten Hacken, Pius & Panocová, Renáta (2022), 'The Suffix *-ation* in English', *Arbeiten aus Anglistik und Amerikanistik* 47(1): 29-57.

OED (2000-2021), *Oxford English Dictionary*, Third edition, edited by John Simpson and Michael Proffitt, Oxford: Oxford University Press, www.oed.com

SNK (2020) *Slovenský národný korpus – prim-9.0-public-all*. Bratislava: Jazykovedný ústav Ľ. Štúra SAV 2020. Available at WWW: http://korpus.juls.savba.sk.