

Muriel Norde, Francesca Masini, Kristel Van Goethem &  
Daniel Ebner

# ***Wannabe* Approximatives**

Creativity, Routinization or Both?

**Abstract:** *Wannabe* is a versatile word that has found its way into several languages. A univerbation of *wanna* (< *want to*) and *be*, English *wannabe* is used both as a noun meaning ‘a fake person’ and as part of a collocation, e.g. *wannabe-gangster* or *Elvis-wannabe*, where it is often depreciative in meaning. This paper presents a corpus-based case study of *wannabe* in English and five other languages *wannabe* has been borrowed into (Danish, Dutch, French, Italian and Finnish). In all languages, *wannabe* collocations are productive and show substantial variation, both between languages and within a single one. This constructional variability of *wannabe* raises interesting questions about creativity and routinization in word formation. We propose an adapted version of the concepts of F- and E-creativity, which distinguishes between F1-, F2- and E-creativity. From the point of view of Construction Morphology, *wannabe* collocations can be formalized as schemas with *wannabe* as a fixed slot and different constraints on the collocate in the open slot. Our case studies show that these schemas are flexible and open to incremental formal or semantic changes (F2-creativity), which may result in either routinization, i.e. adaptation of existing schemas (F1-creativity), or the entrenchment of new ones (E-creativity).

## **1 Introduction**

*Wannabe*, from the phrase *want to be*, is a versatile word that can be used either as a free form or in combination with other words, both in English and in languages it has been borrowed into. As a free form, it mostly occurs as a noun referring to a disingenuous person, as in (1), or, more rarely, as a predicative adjective, as in (2). Collocates include common nouns, as in (3), proper names as in (4), or adjectives as in (5).<sup>1</sup>

---

<sup>1</sup> The examples are from enTenTen20. See Section 4 for a description of the corpora used for this study. All examples are quoted as they occur in the corpus, that is, we did not remove typos or other errors. Also, we do not provide glosses or word-by-word translations, as these are not necessary for this phenomenon.

- (1) *My pal Bruno the Singing DJ is a welcome exception. I hope it's him and not some **wannabe**.*
- (2) *I don't think she's "**wannabe**" at all she's still unique . . .*
- (3) *This drama involves his friend Sumire, an aspiring writer and **wannabe beatnik**.*
- (4) *"The world is full of **Madonna wannabes**, I might have even dated a couple," said Britney Spears' ex. "But there is truly only one Madonna."*
- (5) *A **wannabe-posh**, soapy, sheer, very smooth (and very synthetic) sort of citrus-floral-musk blend*

While the original meaning of 'to aspire to' persists, the implication is often one of ambitions doomed to fail – try as they might, the *Madonna wannabes* in example (4) can never compare to the real one. This mildly depreciative sense opened the door to more general, often approximative, meanings, as in the adjectival phrase in (5), but also in combination with nouns that do not refer to a human being, such as *BMW wannabe* or *wannabe bestseller*. Examples (1–5) furthermore show that collocational *wannabe* is quite variable: it occurs both with a hyphen or without, and it can either precede its collocate or follow it. More variation still is found in other languages, where *wannabe* can form part of proper compound words, for instance Dutch *wannabepiloot* 'wannabe pilot' or Finnish *wannabekirjailijat* 'wannabe authors'.

In this paper, we present data from contemporary web corpora to compare the various uses of *wannabe* in English and five recipient languages (Danish, Dutch, French, Italian and Finnish). For each language, we analyse the position of *wannabe* in relation to its collocate, its (orthographical) bonding properties, the part of speech of the collocate, frequency distributions, and semantic categories of the collocate. Our analysis shows substantial language-internal variation but also differences between the languages in our set, both in terms of the constructional properties just mentioned and the degree to which the recipient languages are similar to English.

For these reasons, our data lend themselves well to questions concerning the concept of creativity, in particular how creativity can be delineated from routine in word formation. Creativity has often been understood as the "intentional manipulation of linguistic structure" (Bergs 2018: 290), whereby the hearer can discriminate between creative use, i.e. breaking *conventions*, and errors, i.e. breaking *rules* (Uhrig 2020: 1). To distinguish between creativity and routine (or

regular, productive usage), we refer to Sampson’s (2016) notions of F-creativity (fixed creativity), defined as “activities which characteristically produce examples drawn from a fixed and known (even if infinitely large) range” and E-creativity (enlarging or extending creativity), i.e. “activities which characteristically produce examples that enlarge our understanding of the range of possible products of the activity” (Sampson 2016: 17). In this dichotomy, F-creativity corresponds to routine, whereas E-creativity corresponds to creativity in Bergs’ sense.

However, the difference is difficult to operationalize because there is a ‘grey area’ where speakers may produce new instances that still follow established patterns of word formation, but which break conventions nevertheless, for instance when they use *wannabe* to modify an adjective or a non-human noun. But when this new usage catches on and type frequency increases, the rule becomes less constrained, such that collocations are no longer restricted to human nouns. Once that has happened, it is no longer justified to speak of breaking conventions. Norde & Trousdale (2024) therefore propose a more fine-grained model which acknowledges two stages of F-creativity. In their model, F1-creativity is used for fully productive patterns (such as de-adjectival nouns in *-ness*), whereas F2-creativity refers to a situation where the old pattern comes to be less constrained (in our case, collocations with an adjective or non-human noun). E-creativity, finally, is the stage when a new rule has been fully established. Since the distinction between the various categories of creativity is gradual rather than absolute, we use the term ‘routinization’ to refer to emerging patterns.

The remainder of this paper is organized as follows: in Section 2 we discuss how *wannabe* is an example of approximative morphology and we present our research questions in Section 3. In Section 4, we explain how we collected our corpus data and annotated them in Excel. After a survey of the quantitative results of our study in Section 5, we discuss the implications of our study for the debate on creativity versus routinization in Section 6, in particular with respect to the concepts of F1-, F2- and E-creativity. In Section 7, finally, we present our conclusions and offer suggestions for future research.

## 2 Approximative Morphology

### 2.1 Modality as a Source of Approximation

APPROXIMATION is a concept well studied in various subfields of linguistics, in particular pragmatics and discourse studies, but only recently has it gained attention within morphology (Masini, Norde and Van Goethem 2023). Following Masini &

Micheli (2020: 384–385), we understand APPROXIMATION as a complex functional domain comprising different (although related) values such as FAKENESS, IMITATION, RESEMBLANCE OR INTENTIONAL VAGUENESS. In Grandi & Körtvélyessy's model of evaluative morphology (2015: 11), APPROXIMATION involves a shift towards the 'negative' pole of evaluation. Examples from word-formation include compounding elements expressing IMITATION/FAKENESS in Dutch (e.g., *kunstgras* 'artificial grass' or *namaak-wasabi* 'fake wasabi', cf. Van Goethem and Norde 2020), prefixal elements like Italian *simil-* (e.g., *simil-marsupio* 'sort of marsupium/pouch', *freddo simil-siberiano* 'Siberian-like cold', cf. Masini and Micheli 2020) as well as suffixes like English *(-)ish*, which is perhaps the most studied approximative morpheme to date (e.g., Oltra-Massuet 2017; Kempf and Eitelmann 2018; Eitelmann, Haugland and Haumann 2020; Eitelmann and Haumann 2023).

From these studies, it emerges that approximative morphemes may develop out of a variety of sources (Masini, Norde and Van Goethem 2023: 7–10) some of which are better recognized and explored than others. Well-studied sources include words meaning 'fake', e.g. Greek *pseudēs* 'false', adopted in many other languages as the prefix *pseudo-* (cf. Vassiliadou et al. 2023; Van Goethem, Norde and Masini 2023); diminutives (Dressler and Merlini Barbaresi 1994; Grandi 2017); degree and quantity items such as *quasi-* in English (e.g., *quasi-particle* cf. Cappelle, Daus and Hartmann 2023) or *semi-* in Italian (e.g., *semi-official*; cf. Micheli 2023); spatial (proximity) items, such as *near-* in English (e.g., *near-synonym*; cf. Cappelle, Daus and Hartmann 2023) or *para-* in Italian (from Greek *para* 'beside'; cf. Micheli 2023); simulative items, such as English *-like* (*prefix-like*) or Italian *simil-* (Masini and Micheli 2020); relational items, such as the already-mentioned English *-ish* or Dutch *-achtig* (*groenachtig* 'green-ish').

A source that, to our knowledge, has not yet been explored in morphological studies on approximation, is modality. Nevertheless, examples are found in various languages, such as epistemic adverbs or volitional expressions. The conceptual step here is from an assessment of speakers' (un)certainity about what they say (epistemic modality) or an aspiration to attain a (still unreal) specific state-of-affairs (volitional modality) to approximation as a process that expresses a qualitative deviation from a standard or default value, towards the 'negative' pole (cf. Grandi and Körtvélyessy's model mentioned above).

As for epistemic adverbs, English *maybe*, for instance, is found in compound-like expressions conveying approximation, combined with both nouns (6) and adjectives (7). The case in (6) is especially interesting since adverbs are not supposed to modify nouns.

- (6) *Not at all, I think these history threads are awesome and want to add my little bits of trivia. I'm just not willing to attach specific names to my admittedly fuzzy memory of the **maybe-incident**.* [enTenTen20]
- (7) *The best of the bunch is alternate history drama “For All Mankind,” from the creator of “Battlestar Galactica,” but is it worth the money to see just one **maybe-good** show?* [enTenTen20]

The same applies to Italian *forse* and French *peut-être* ‘maybe, perhaps’, which are found in combination with nouns (mainly) (8), adjectives (9) and even with full phrases (10) (examples from itTenTen20 and frTenTen20):

- (8) [Italian] *Sorella di mezzo ha diciott'anni, una sfilza di fratelli e sorelle, una madre bigotta che la vorrebbe già sposata, il ricordo di un padre morto di depressione, e un **forse-fidanzato**.* ‘Middle Sister is eighteen years old, a string of siblings, a bigoted mother who would already like her married, the memory of a father who died of depression, and a **maybe-boyfriend**.’
- (9) [Italian] *Da oggi dunque una città ricoperta di rumenta non sarà più colpa di scelte sbagliate, di strategie folli, di priorità errate e di impostazioni politico-ideologiche malate, no, sarà tutta colpa del misterioso incendio **forse-doloso**.* ‘Starting today, therefore, a city covered in rubbish will no longer be the fault of bad choices, foolish strategies, wrong priorities, and sick political-ideological approaches, no, it will be all the fault of the mysterious **maybe-fraudulent** fire’.
- (10) [French] *Typiquement le message qui fera fuir **un peut-être futur utilisateur de Linux** !*  
‘Typically the message that will scare off a **maybe-future Linux user**!’

Volitional expressions, too, can be employed as approximating strategies, often carrying a pejorative meaning derived from the ‘aspiring’ semantics, which implies the non-realization of a desired state. Examples are German *möchtegern* ‘would very much like to’ in (11), Italian *vorrei-ma-non-posso* in (12) (a lexicalized clause – literally meaning ‘I would like but I can’t’ – used to refer, often ironically, to some aspiration impossible to achieve), and of course English *wannabe* (as in *wannabe popstar*). This *wannabe* construction is the case-study that we are going to discuss in detail in the following sections.

- (11) [German] *Die Mehrzahl der Manuskripte fristete ein Schattendasein in den Schubladen frustrierter **Möchtegern-Schriftsteller**. ‘Most of the manuscripts were leading a shadow existence in the drawers of frustrated **wannabe-authors**’ [deTenTen20]*
- (12) [Italian] *In questi giorni di mare e focaccia, [. . .] ho rivisto uno di questi teli utilizzato per coprire la vetrina di una **boutique vorrei-ma-non-posso** di paese, sempre in Liguria ovviamente. ‘In these days of sea and flatbread, [. . .] I saw again one of these cloths used to cover the window of a village **wannabe boutique**, again in Liguria of course.’ [itTenTen20]*

## 2.2 English *wannabe*

According to the OED ([www.oed.com](http://www.oed.com)), *wannabe* is a slang (originally U.S.) word created by the univerbation of *wanna* (itself a univerbation of *want* and *to*) and *be* (cf. also earlier *would-be*, used as an adjective). *Wannabe* is registered:

- (i) as a noun<sup>2</sup> (“A person who tries to emulate someone else, esp. a celebrity, in appearance and behaviour; a person who wants to belong to and tries to fit in with a particular group of people. Frequently *depreciative*”), when used either independently (13) or after another noun (14);
- (ii) and as an adjective<sup>3</sup> (“Desiring or aspiring to be a specified person or type of person; would-be. Also in extended use of a product designed to emulate or rival another. Sometimes *depreciative*”), when preceding another noun (15).
- (13) *Many at Cambridge will tell you that the drama world there is filled with ambitious, pretentious, bitchy **wannabes**. [OED: S. Fry, Fry Chronicles 95, 2010]*
- (14) *A morbid Madonna-**wannabe** fascinated with tabloid tales of bizarre deaths. [OED: Washington Post 17 January 1986 (Weekend section) 8/4]*
- (15) *There was little doubt that Bill Gates was intent on building a **wannabe** Mac for the vast PC market. [OED: Rolling Stone 18 April 1996 61/1]*

<sup>2</sup> Oxford English Dictionary, s.v. “wannabe, n.”, July 2023 (<https://doi.org/10.1093/OED/9474186542>).

<sup>3</sup> Oxford English Dictionary, s.v. “wannabe, adj.”, July 2023 (<https://doi.org/10.1093/OED/3277987720>).

This fluid categorial status tallies with Brinton & Traugott's (2005: 16) gradient view on categories, which is considered as “a factor motivating change, and also as the outcome of changes in usage”.

According to the OED, the semantics of *wannabe* is often depreciative (equivalent to “phoney” or “fake”) when used as a noun (cfr. (13)) and only “[s]ometimes *depreciative*” when used as an adjective. The first use of *wannabe* recorded in the OED traces back to 1976:

- (16) *At 38 she had 21 years of racket life behind her. Whereas Joe, that year, was still a Jimmy Cagney wannabe.* [OED: New York Magazine 26 July 1976 43/3]

Crucially, unverbated *wannabe* can still function as a verbal phrase, as illustrated in the following example, in which the second and third occurrences of *wannabe* function as verbal phrases syntactically:

- (17) *Our children are not wannabe gangsters, they wannabe loved for who they are, and they wannabe great!* [enTenTen20]

### 3 Research Questions

In this study, we explore various construction types involving *wannabe* in six languages (English, Dutch, Danish, French, Italian and Finnish), with the aim to answer the following three research questions:

**RQ1:** How are the different construction types with *wannabe* distributed across each language?

**RQ2:** What are the formal and semantic properties of *wannabe* collocations?

**RQ3:** How creative are *wannabe* collocations?

We look at *wannabe* both as a free noun and in other constructions, such as combinations with other nouns or predicative usage. Our first research question therefore refers to the distribution across the six languages in our dataset. More specifically, we are interested in the collocational behaviour of *wannabe* cross-linguistically. As combinations of words are construed differently in the languages in our set (three Germanic languages, two Romance languages and one Finno-Ugric language), we avoid the term ‘compound’. As we explain in more de-

tail in Section 4, the term ‘collocation’ may refer to compounds written as one word, such as Danish *wannabehjemmeside* ‘wannabe homepage’, but also to expressions such as English *wannabe starlet*<sup>4</sup> or French *écrivaine wannabe* ‘wannabe author’. Our second research question, then, pertains to formal and semantic properties of *wannabe* collocations of all types. Our third and final research question relates to the topic of this collective volume and our three-way model of creativity outlined in Section 1. For each language in our sample, we consider the various construction types found in that language, in order to assess whether these patterns are F1-, F2- or E-creative.

## 4 Data and Methods

To answer the aforementioned research questions, we conducted a corpus-based study of *wannabe* formations in English, Danish, Dutch, French, Italian and Finnish. The CQL query [word = ".\*wannab.\*"] allowed us to extract formations starting or ending with the sequence *wannab*, including different spelling variants such as *wannabe(s)*, *wannabee(s)* or *wannabie(s)*. We downloaded random samples of 1000 tokens per language from the TenTen web corpora at Sketch Engine (Kilgariff et al. 2014).<sup>5</sup> These samples were then manually checked, and irrelevant occurrences were removed until we obtained 500 relevant occurrences per language. Occurrences considered irrelevant included metalinguistic reference to *wannabe*, references to the Spice Girls’ *Wannabe* song, *wannabe* used in URLs and account names (e.g. *wannabe\_god*), other strings containing *wannab*\* (e.g. *wanna-bike.com*), literally repeated sentences, and *wannabe* used in entire English sentences in the non-English datasets.

We then annotated those 500 occurrences for a diverse set of orthographic, morphological and semantic properties.

First, we annotated the **construction type** and distinguished the following categories:

---

<sup>4</sup> English compounding differs from compounding in other Germanic languages in that English compounds may be written as one word (*oatcake*) or two (*cherry cake*) and it is a moot point whether multiword constructions are morphological or syntactic constructions. In spite of extensive literature on the topic, there appear to be no decisive criteria, like primary stress or compositional semantics (see Bauer 2019 for a review). See Section 4 for further discussion on terminology.

<sup>5</sup> More specifically, we used the following corpora: enTenTen20 for English, daTenTen20 for Danish, nlTenTen20 for Dutch, frTenTen20 for French, itTenTen20 for Italian and fiTenTen14 for Finnish.



(i) **Wannabe used as a noun** (18–19):

- (18) [Dutch] *Anders denken mensen dat je niet meer bent dan een wannabe.* ‘Otherwise, people will think you are nothing but a **wannabe**.’
- (19) [Danish] *Resultatet er, at ingen er særlige, men at vi alle bare er kopier og wannabes.* ‘As a result, nobody is special, we are all just copies and **wannabes**.’

(ii) **Wannabe used as an adjective**, either predicative (20) or attributive. To distinguish between attributive use and the “collocation” type (see below), adjectival use was restricted to unambiguous cases including coordination with another adjective (21), modification by a degree adverb (22), or specific adjectival constructions, for instance [*un N de ADJ*] in French (23).

- (20) [Danish] *Det er flot og anderledes på den fede måde – ikke bare wannabe.* ‘It is nice and different in a cool way, not just **wannabe**.’
- (21) [English] *No disagreement on the point about actual and wannabe plutocrats either (. . .)*
- (22) [Danish] *Hvis man har oplevet La Santa, så forekommer Club Med (som jeg også har boet på en enkelt gang) som ren wannabe.* ‘If you have experienced La Santa, then Club Med (where I also stayed a couple of times) appears **totally wannabe**.’
- (23) [French] *Ceci est un film d’action de wannabe (. . .)* ‘This is a **wannabe action movie** (. . .)’

(iii) **Wannabe used as an ‘adverb’ or discourse marker** (24):

- (24) [Italian] *Seavessi non scrive per paura della commessa. Su Wikipedia alla voce wannabe, c’è la foto di Seavessi. Voi cosa wannabe vorreste?.* ‘Seavessi [proper name from sentence ‘if I had’] does not write out of fear of the clerk. On Wikipedia under wannabe, there is a picture of Seavessi. What **wannabe** would you guys like?’

(iv) **Wannabe as part of a “collocation”** (25–26). As already explained in Section 3, we preferred the term ‘collocation’ to ‘compound’ because the dividing line between morphological and syntactic formations is rather difficult to draw, even more so in written data that do not allow us to check the prosodic properties of the sequences. Moreover, the spelling criterion is not reliable either to identify true compounding, especially when combining an English loanword (*wannabe*) with native words.

- (25) [Danish] *Måske wannabe -‘vampyren’ vil stikke sig selv en metaforisk pæl i hjertet (. . .)?* ‘Maybe **the wannabe-‘vampire’** wants to stick a metaphorical stake into his own heart?’
- (26) [French] *Là par contre, j’aurais accepté sans rechigner que tu me traite de “pseudo-Charles Bronson du dimanche” ou de “wannabe-Zorro de l’Internet” ou un truc comme ça (. . .)* ‘On the contrary, I wouldn’t have minded you calling me a “a Sunday pseudo-Charles Bronson” or a “**wannabe-Zorro** of the Internet” or something like that (. . .)’

(v) **Derivations of wannabe** (27):

- (27) [Dutch] *Dit is wel keihard afrekenen met de wannabe-ers a.k.a privé lease-rijders.* ‘This does deal harshly with the **wannabe-ers** a.k.a. private lease drivers.’

For the ‘**collocation**’ type, which is the main focus of this study, we annotated some additional properties.

- The first one is **orthographic bonding of the collocations**, which distinguishes between free (28), bound (29) and hyphenated (30) collocations, or a ‘hybrid’ combination of some of the former spelling strategies (31).
- (28) [French] *Soufron, vous n’êtes même pas un bobo, vous êtes ce que j’appelle un **wannabe bobo**.* ‘Soufron, you’re not even a bobo, you’re what I call a **wannabe bobo**.’
- (29) [Dutch] *Ook beschikte de 37-jarige **wannabepiloot** niet over een vliegbrevet.* ‘Also, the 37-year-old **wannabe pilot** did not have a pilot’s license.’

- (30) [Danish] *Musikken lyder måske mest som hvis nu Robbie Williams' sangskriverteam havde skrevet en **wannabe-Oasis-sang***. 'The music sounds perhaps mostly as if Robbie Williams' song-writing team had written a **wannabe-Oasis song**.'
- (31) [Danish] *Jeg synes, vi skal droppe det der "**nationalsangs wannabee**" fuldstændig*. 'In my view, we should get rid of this "**national anthem wannabe**"'
- Second, we annotated the **part of speech (POS) of the collocate**, as well as its **lemma form** to facilitate productivity calculations (number of types and hapaxes). The following examples illustrate collocates in the form of a (proper) noun (32), a noun phrase (33), an adjective (34) and a verb phrase (35).
- (32) [Danish] *Sproget – en såkaldt "indre monolog" – lyder, som var det skrevet af en **James Joyce-wannabe***. 'The language – a so-called 'inner monologue', sounds as if it had been written by a **James Joyce wannabe**.'
- (33) [Dutch] *Een aanrader voor elke **wannabe journalistieke detective*** 'Highly recommended for any **wannabe journalistic detective**'
- (34) [Finnish] *Nonniin se niistä mysteerisistä **wannabe-aistikkaista** ajatuksista*. 'Well, so much for those mysterious, **wannabe-tasteful** thoughts.'
- (35) [Dutch] *(. . .) hij wordt een dagje ouder en T is natuurlijk volop in zijn **wannabe man zijn periode*** '( . . .) he's getting a day older and is obviously fully into his **wannabe manhood** [lit. man be] period'
- Third, we annotated the **word order** of the collocations, distinguishing between three schemas: [X-wannabe], [wannabe-X] or [X-wannabe-X]. For the first two options, see for instance the difference between (32) and (33). We find the third pattern in (36).
- (36) [Italian] *Kiefer Sutherland non mi è mai stato troppo simpatico, probabilmente perchè in questo film interpretava un **vampiro-wannabe-biker***, (. . .). 'I've never liked Kiefer Sutherland too much, probably because in this film he played a **vampire-wannabe-biker**, (. . .).'
- A final formal parameter included in the annotations is the **presence of inflection on wannabe**. We find both native and borrowed (= English) inflec-

tion: see (37) for number and case inflection in Finnish, (38) for English plural inflection for number in Italian. However, since the vast majority of results are in the (uninflected) singular, we decided not to include this parameter in our quantitative analysis.

- (37) [Finnish] (. . .) *me täällä ollaan niiden rinnalla ihan **wannabeita***. ‘We here are just **wannabes** compared to them.’
- (38) [Italian] *questi **wannabees** possono essere noiosi* ‘these **wannabees** can be boring.’

When *wannabe* collocates with a noun, we distinguished three **semantic categories**: HUMAN to refer to human beings (living, dead, or fictional) or groups of human beings, such as rock bands, as in (39); ANIMATE to refer to all living creatures that are not human (animals and fantasy figures such as vampires or zombies) as in (40); and INANIMATE to refer to objects, food, abstract concepts etc., as illustrated in (41).

- (39) [French] *La vidéo montre des mineurs, parfois adolescents, défiler devant un parterre de **wannabe parents***. ‘The video shows minors, sometimes teenagers, parading in front of an audience of **wannabe parents**.’
- (40) [Danish] *den får de to af monsteret/ **wannabe zombien** der kommer i slutningen* ‘it gets the two of them from the monster/ **wannabe zombie** that comes at the end’
- (41) [Dutch] *Gember heeft een eigenschap die het ver boven alle **gember-wannabe** middeltjes plaatst*. ‘Ginger has a property that places it far above all **ginger-wannabe** remedies.’

We now turn to the results section, where the focus will be on the collocates in the ‘collocation’ type.

## 5 Results

In this section, we present the results of the case studies, starting with the different construction types in which *wannabe* is used and then focusing on the formal and semantic properties of a specific construction: *wannabe* collocations.

### 5.1 Construction Types

**Table 1:** Construction types.

Construction type	English	Danish	Dutch	French	Italian	Finnish
collocation	357	342	370	294	340	393
collocation-derivation	1	7	2	0	0	2
derivation	0	4	2	6	2	1
embedded collocation	4	30	11	3	4	34
predicative	2	5	12	6	4	19
wannabe_ADJ	3	1	3	9	12	1
wannabe_N	132	111	100	182	135	50
wannabe_V	0	0	0	0	2	0
wannabe_ADV	0	0	0	0	1	0
wannabe_VP	1	0	0	0	0	0
<b>Total</b>	<b>500</b>	<b>500</b>	<b>500</b>	<b>500</b>	<b>500</b>	<b>500</b>

As shown in Table 1, *wannabe* is used in a variety of construction types in the data samples, but frequencies differ from sample to sample. Collocations with *wannabe*, as in (42), form the biggest part of all language samples (a more detailed analysis of collocational constructions follows in sections 5.2 and 5.3).

(42) [English] *I am a wannabe rocker trapped in folky singer-songwriter schtick.*

Apart from “simple” collocations, all languages also feature the use of *wannabe* in embedded collocations, such as (43), where the collocation *Elvis Presley wannabe* is used to modify the noun *frisure* ‘haircut’. Such embedded collocations were found predominantly in Finnish and Danish (and to a certain degree in Dutch) and point to a more synthetic use of *wannabe*.

(43) [Danish] (. . .) *og derfor er det ikke så underligt, at Cage denne gang er ude i en Elvis Presley wannabe-frisure.* ‘and therefore it is not so remarkable that Cage this time is sporting an Elvis Presley-wannabe-haircut.’

(44) [Finnish] *Tämä on sellainen wannabe-feather and fan-kaulaliina* (. . .) ‘This is such a wannabe-feather and fan-scarf (. . .)’

The Romance languages, by contrast, show more analytic uses of *wannabe* as independent noun (45) or adjective (46).

- (45) [Italian] *Ormai, il web è divenuto solo una vetrina per wannabe piuttosto sfigati o – nel migliore dei casi – un mezzo a disposizione di chi già scrive.* ‘By now, the web has become only a showcase for rather lame **wannabes** or – at best – a medium available to those who already write.’
- (46) [French] *D’entrée de jeu, oubliez tout de suite la selfie prise à bout de bras dans le salon. Ça fait très wannabe et vous ne voulez pas ça!* ‘Forget about the arm’s length selfie in the living room. It looks very **wannabe** and you don’t want that!’

Whereas the English dataset almost exclusively consists of the use of *wannabe* in collocations and in nominal use, a more varied range of construction types can be observed in the recipient languages. These include the use of *wannabe* as a verb (47) and the use of *wannabe* or a *wannabe*-collocation as bases for derivation (48–49).

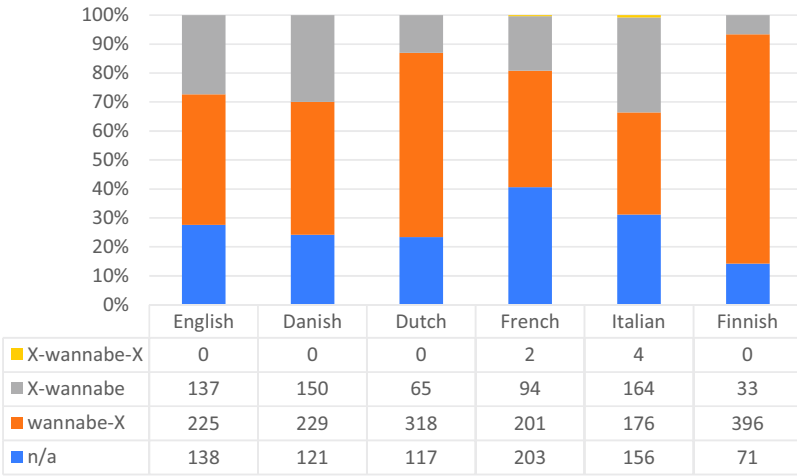
- (47) [Italian] *Detto questo:-D io wannaberei . . . dire al mio capo (e a mio marito) quanto mi faccia schifo fare il mio lavoro e che mi licenzio* ‘That said:-D I **would like to** . . . tell my boss (and my husband) how much I hate my job and that I’m quitting.’<sup>6</sup>
- (48) [Danish] *Nogle synes at det helt vildt sjovt at bruge engelske låneord, andre at det lyder wannabeagtigt.* ‘Some think it is totally cool to use English loanwords, others think it sound **wannabe-like**.’
- (49) [Finnish] *Kirjoittajan artikkeli sinänsä on minun mielestäni suurelta osin tyhjänpäiväistä carriebradshaw-wannabe-maista läpinää, enkä muuten halua sitä kommentoida [. . .]* ‘The author’s article itself is, in my opinion, largely empty **Carriebradshaw-wannabe-like** nonsense, and I don’t want to comment on it [. . .]’

The use of *wannabe* in a more varied range of constructions can be seen as creative extensions of *wannabe* word formation patterns in the recipient languages; however, the frequencies of construction types only found in the recipient languages (but not in English) are very low.

<sup>6</sup> This example is peculiar because *wannabe* is used to mean *would like (to)*, with the inflectional suffix *-erei* (1.SG.COND) attached to it, marking its verbal status.

## 5.2 Formal Properties of Wannabe Collocations

### 5.2.1 Word Order



**Figure 1:** Word order.

Figure 1<sup>7</sup> illustrates the proportions of the three word-order patterns. Since *wannabe* acts as an approximative modifier in the collocation type, we would expect to find [*wannabe*-X] order in the Germanic languages, because of their general Modifier-Head structure, and the opposite order [X-*wannabe*] in the Romance languages, which are characterized by Head-Modifier structure.

With respect to Germanic (English, Danish, Dutch), we indeed notice a preference for [*wannabe*-X] order, but, especially in English and Danish, also a surprisingly high proportion of the inverse word order (see for instance (50)). The status of *wannabe* in such instances is not crystal-clear: does it function as the head of the sequence or as an approximative marker? This issue needs further investigation.

(50) [English] *Some have described Departed as feeling like a Scorsese-wannabe created it*

<sup>7</sup> In this Figure and elsewhere, “n/a” stands for “not applicable”; in this case, it refers to *wannabe* constructions that are not collocations and hence do not have word order.

The [*wannabe*-X] order is clearly predominant in Finnish, as illustrated in (51).

- (51) [Finnish] (*. . .*) *onko tää joku median neuvo wannabejulkiksille että jos tahdotte saada huomiota niin provosoikaa mahdollisimman paljon.* ‘(*. . .*) is this some media advice to **wannabe celebrities** that if you want to get attention, provoke as much as possible’

Whereas the inverse order [X-*wannabe*] was expected in the Romance languages, Italian and especially French seem to show a high proportion of ‘pattern borrowing’ (Gardani 2020) of the English word order pattern. Moreover, contrasting examples like (52–53) suggest that the word order pattern is quite arbitrary in French.

- (52) [French] *Il n’est plus un Vader-wannabe, il est Kylo Ren.* ‘He is no longer a **Vader-wannabe**, he is Kylo Ren.’
- (53) [French] *Kylo devient définitivement le wannabe Vader Fucky Ryan* ‘Kylo definitely becomes the **wannabe Vader** Fucky Ryan’

[X-*wannabe*-X] is very rare, it is only found a couple of times in the French and Italian datasets. An example is given in (36) above.

### 5.2.2 Orthographic Bonding

With respect to orthographic bonding, we do not find clear routinization in this field, but some general tendencies can be detected. Overall, we notice more free uses in English and Romance (see for instance *wannabe rocker* (42) for English and *wannabe bobo* (28) for French) and more bound (especially hyphenated) cases in the other Germanic languages and Finnish (see for instance *wannabepiloot* ‘wannabe pilot’ (29) for Dutch and *wannabe-aistikkaista* ‘wannabe tasteful’ (34) for Finnish).

These results are in line with the assumption that Romance languages are more analytic than Germanic languages and that, within the Germanic language family, English is situated at the analytical end of the cline (see, e.g., Van Haerlingen 1956; Hüning et al. 2006; Lamiroy 2011; König and Gast 2018; Van Goethem, Norde and Masini 2025).



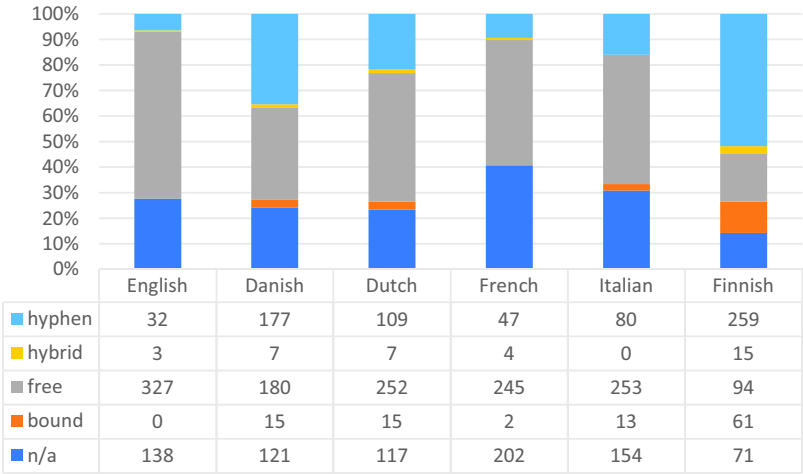


Figure 2: Orthographic bonding.

5.2.3 Part of speech of the collocate

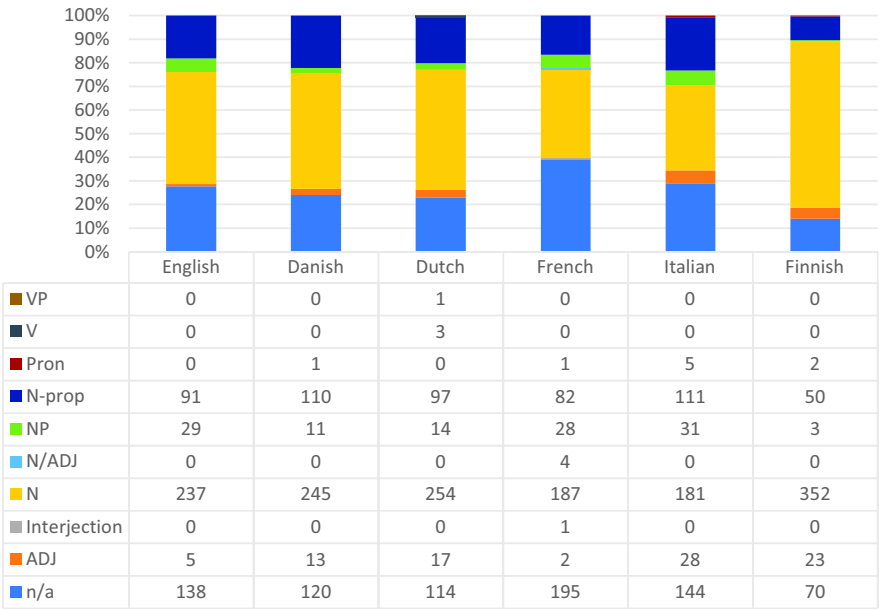


Figure 3: Part of speech of the collocate.

With respect to the part of speech of the collocate, a first observation is that in English *wannabe* almost exclusively combines with nominal collocates (N, Proper N, NP), as in (54).

- (54) [English] *Is ‘whole hearty wheat’ really a burden to good health and **wannabe beach bodies**?*

However, as Figure 3 shows, in the recipient languages we also find creative expansion to other categories, such as adjectives (55), pronouns (56) and verbs (57).

- (55) [Italian] *Gli attori dietro gli imbarazzanti personaggi sono altrettanto allucinati, specialmente considerando una trama **wannabe impegnata** con risvolti drammatici nelle vite di ognuno.* ‘The actors behind the awkward characters are equally terrible, especially considering a **wannabe committed** plot with dramatic implications in everyone’s lives’
- (56) [Italian] *Da anni ormai infatti la kermesse cinematografica fa da vetrina a una pletora di attori, cantanti, celebrità e **wannabe tali**, (. . .).* ‘For years now, in fact, the film festival has served as a showcase for a plethora of actors, singers, celebrities and **wannabe such**, (. . .)’
- (57) [Dutch] *Ja, dat noemen we **wannabe modden**.* ‘Yes, that is what we call **wannabe modding** (moderating).’

Moreover, we find routinization of proper names in all languages (for instance *Scorsese-wannabe* in (50)), if to a lesser degree in Finnish. In Section 5.3, we will have a closer look at the semantics of these proper names.

#### 5.2.4 Productivity

In this section, we only consider the productivity of collocations with common nouns.<sup>8</sup> We consider *wannabe* collocations to be schemas with one fixed slot and

---

<sup>8</sup> As we demonstrated in Section 5.1, many collocations involve proper names. We did not consider these constructions in our productivity measures, because proper names have unique reference, i.e. they describe specific individuals and properties that are associated with them (Ainiala and Östman 2017: 6; Kosse 2021: 1). Adding them to the slot filler types would mean a substantial increase in the number of hapaxes, resulting in a higher potential productivity than is warranted for this schema.

one open slot and we use productivity measures to compare the lexical diversity of the slot fillers in the different language samples. As the figures in Table 2 show, the vast majority of collocations consists of hapaxes (i.e. types that occur only once in the sample). Collocations with a token frequency higher than 1 are rare and have maximum frequencies between 4 and 6 for most languages. Only Finnish has one type (*kirjalija* ‘author’) with a slightly higher token frequency of 17.

The large proportion of hapaxes results in high productivity ratios, not just the Type / Token Ratio (TTR), but also the Hapax / Token Ratio, or Potential Productivity (PP) and the Hapax / Type ratio (HTR). ‘Potential Productivity’ (Baayen 2009) can be interpreted as an indication of (potential) vocabulary growth. Especially in large corpora, hapaxes can be seen as a proxy for neologisms (Zeldes 2013: 61), so that a high Hapax / Type ratio suggests that a pattern is relatively new with great potential to expand. The figures in Table 2 suggest high productivity of *wannabe* collocations in all languages.

Table 2: Productivity.

Productivity	English	Danish	Dutch	French	Italian	Finnish
TTR	318/362 = <b>0.88</b>	336/380 = <b>0.88</b>	330/390 = <b>0.85</b>	252/306 = <b>0.82</b>	282/340 = <b>0.82</b>	345/430 = <b>0.80</b>
PP	291/362 = <b>0.80</b>	303/380 = <b>0.80</b>	292/390 = <b>0.75</b>	219/306 = <b>0.71</b>	253/340 = <b>0.74</b>	305/430 = <b>0.71</b>
HTR	291/318 = <b>0.92</b>	303/336 = <b>0.90</b>	292/330 = <b>0.88</b>	219/252 = <b>0.87</b>	253/282 = <b>0.89</b>	305/345 = <b>0.88</b>

### 5.3 Semantic Category of the Collocate

The last property we investigated is the semantic category of the collocate. As illustrated in Section 4, we used a rather coarse-grained classification, distinguishing three semantic classes: HUMAN, ANIMATE, and INANIMATE.

As we can see in Figure 4, collocates in English belong predominantly to the human category (cf. (58)), whereas the proportion of human collocates in the recipient languages is lower: except French, all other languages display a creative extension to inanimate collocates (animate ones are very marginal everywhere), which is more marked in Danish and Italian, and slightly less prominent in Finnish and Dutch.

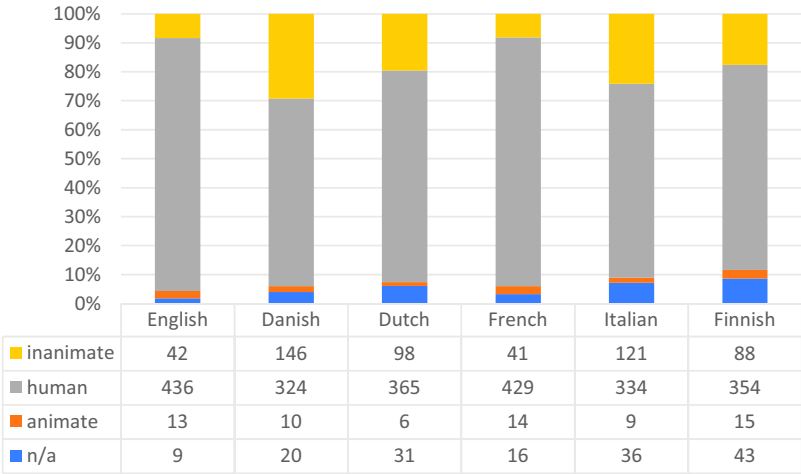


Figure 4: Semantic category of the collocate.

(58) *But I don't agree with the suggestion that every wannabe-blogger must opt for Aweber.*

The combination with inanimate entities is an indication of the change *wannabe* is experiencing in these contexts: the original ‘aspiring’ meaning (cf. (58)), hinting to a possible future state-of-affairs to attain, is not compatible with inanimate entities that, being non-sentient, cannot aspire to anything, strictly speaking. This triggers a more abstract interpretation of ‘aspiration’ as deviation from a norm or standard, hence the approximative semantics. The contrast can be seen in examples (59–60), which contain the same expression (*wannabe actress*): in (59), Jonni Kendall is not an actress but an aspiring one, whereas in (60) Katrina is an actress (since she is indeed already acting), but not a full-fledged, famous one (because she is acting in sleazy films). This is also evidence that the approximative meaning applies to human nouns too.

(59) [ . . . ] *Jonni Kendall, a nineteen-year old wannabe actress.*

(60) [ . . . ] *‘Sheila Ki Jawani’ was supposed to be a raunchy item Bhojpuri kind of song as Katrina’s a wannabe actress acting in sleazy films [ . . . ]*

Whereas human collocates are typically common nouns referring to (desirable) occupations (e.g., influencer, actress, blogger, author, etc.) or proper names of celebrities (e.g. Madonna or Elvis as exemplified above), inanimate collocates be-

long to a variety of semantic subcategories, some of which are however recurring, namely: food items (61), or inanimate proper names such as (video) games (62), movie and series titles (63), or brand names (phones, cars, fashion houses, food chains, etc.) (64–65). These inanimate collocations suggest the emergence of specific semantic clusters, which we discuss in Section 6.

- (61) [Danish] *Total bagedag, nu også med **marmorkage-wannabe***. ‘A full baking day, now including a **marble cake wannabe**.’
- (62) [Italian] *Journey è un **mario64 wannabe**, si salta in giro però è lento, scomodo e non succede una cippa* ‘Journey is a **mario64 wannabe**, you jump around but it is slow, uncomfortable and nothing happens’
- (63) [French] *voici un article qui tente d’expliquer le succès d’audience de Revolution, comparée aux autres **Lost-wannabes*** ‘here’s an article that tries to explain Revolution’s audience success, compared to other **Lost-wannabes**’
- (64) [Italian] *La moda costa. Magari non prendevo Gucci, ma meglio H&M di un **wannabe Vuitton***. ‘Fashion costs money. Maybe I wouldn’t get Gucci, but better H&M than a **wannabe Vuitton**.’
- (65) [Dutch] *We sluiten af met een bezoek aan een **wannabee Macdonalds**, Jollibee, hier the place to be voor iedereen, waar ze overheerlijke spaghetti met friet verkopen (. . .)* ‘We conclude with a visit to a **wannabe Macdonalds**, Jollibee, here the place to be for everyone, where they sell delicious spaghetti with fries (. . .)’

It is perhaps not surprising to find many inanimate proper names in our data: famous brands and popular products are easily interpretable as something to aspire to (just like celebrities), which explains why *wannabe*’s creative expansion is going in this direction. However, we also find examples that don’t comply with this, like (66–67), which better exemplify the approximative semantics of *wannabe*.

- (66) [Dutch] *Een zweetdruppel loopt langs mijn **wannabe-bakkebaarden** en een andere teistert mijn linkerwenkbrauw (. . .)* ‘A drop of sweat runs down my **wannabe sideburns** and another plagues my left eyebrow (. . .)’

- (67) [Finnish] *Kun ei parempaa ole, niin olen sitten vanhoja joulukortteja koonnut wannabe-kuuseksi oveen.* ‘For want of anything better, I assembled old Christmas Cards into a **wannabe-spruce** on the door.’

## 5.4 Summary and Analysis

In this section, we address our first two research questions and summarize our findings from the preceding sections, concluding with a statistical analysis of the differences between English and each of the recipient languages. The first two RQs are repeated here for convenience:

**RQ1:** How are the different construction types with *wannabe* distributed across each language?

**RQ2:** What are the formal and semantic properties of *wannabe* collocations?

Overall, our comparative analysis brought to light substantial variation in *wannabe* constructions. Collocations prevail in all languages, ranging from 294 tokens in French to 393 tokens in Finnish. Conversely, *wannabe* as an independent noun is most common in French (182 tokens) and least in Finnish (50 tokens). When we look at the semantic categories of *wannabe* constructions (both of *wannabe* as a free noun and *wannabe* collocations), HUMAN is the most common category, although the proportion is higher in English than in the recipient languages (except French).

Turning now to *wannabe* collocations, we examined five properties: word order, type of (orthographic) bonding, part of speech of the collocate, productivity, and semantic category. Word order, it turned out, differs from language to language, with a strong preference for *wannabe*-X in Dutch and Finnish, a less pronounced preference in English, Danish and French, and an almost 50/50 distribution in Italian. Where type of bonding is concerned, we found substantial differences – collocations are overwhelmingly non-bound in English, Dutch, French and Italian, whereas Danish and, especially, Finnish prefer hyphens in *wannabe* collocations. The part of speech of the collocate is less varied – nouns form the majority, with proper names coming second and adjectives third. Adjectives are however less common in English than in the recipient languages (except French). Differences are also relatively small in productivity scores, which were invariably high, both in terms of Type Token Ratio, Hapax Token Ratio (Potential Productivity) and Hapax Type Ratio. The high number of hapaxes in particular suggests

that *wannabe* collocations are very productive in all languages, with as yet little entrenchment of specific types.

To conclude our quantitative analysis, we compared the recipient languages to English by performing a Fisher's exact test for all formal and semantic properties discussed in Sections 5.2 and 5.3. The results are given in Table 3, where each cell gives the result of a pairwise comparison of a specific variable in English and one of the other languages.<sup>9</sup> We note that for most of these comparisons the difference is significant, with the exception of French, where we did not find statistically significant differences for part of speech, word order and semantic category.<sup>10</sup>

**Table 3:** The recipient languages compared to English.

	Danish	Dutch	French	Italian	Finnish
<b>Construction type</b>	***	** / *	*** / **	ns	***
<b>Orthographical bonding</b>	***	***	** / *	***	***
<b>Collocations: Part of Speech</b>	** / *	** / *	ns	***	***
<b>Collocations: Word order</b>	ns	***	ns	** / *	***
<b>Collocations: Semantic category</b>	***	***	ns	***	***

## 6 Discussion: Creativity or Routinization?

In this section, we return to the concepts of F1-, F2- and E-creativity introduced in Section 1, in order to answer our third and final research question:

**RQ3:** How creative are *wannabe* collocations?

<sup>9</sup> The asterisks in the p-value refer to the significance level (\*:  $p < 0.05$ ; \*\*:  $p < 0.01$ ; \*\*\*:  $p < 0.001$ ), ns means non-significant ( $p > 0.05$ ). For these calculations, we only considered values that were found in all languages, disregarding very infrequent patterns (e.g. *wannabe* as an adverb) that had zeros in most cells.

<sup>10</sup> One reviewer suggests that we do a Bonferroni correction on the significance tests to avoid Type I errors (finding an effect when in fact there is none), arguing that we are doing various tests on the same data. The Bonferroni correction is however not uncontroversial (Cabin & Mitchell 2000) and moreover each cell in Table 3 represents a different cross-tabulation. For these reasons, we are not sure whether the correction is necessary, but we did add the values for the Bonferroni corrections nevertheless. In our case, this meant dividing the significance thresholds by 5 (the number of variables in Table 3), which means the following significance levels: \*:  $p < 0.01$ ; \*\*:  $p < 0.002$ ; \*\*\*:  $p < 0.0002$ . If the resulting significance level changes as a result of this (which in most cases it did not), the adapted value is given after the /.

To account for the patterns we found in our data, we adopt a constructionist approach to morphology (Booij 2010; Booij 2016; Masini & Audring 2019) and morphological change (Norde & Trousdale 2023). In Construction Morphology, complex words “are not seen primarily as a concatenation of morphemes, but as independent meaningful units within which certain subcomponents (morphemes) may be distinguished on the basis of paradigmatic relations with other words” (Booij 2018: 4–5). Generalizations over complex words take the form of schemas, which express the symbolic relation between the formal and the semantic pole of a morphological construction. Schemas are often partially schematic, with an open slot and a filled one. Two constructions with *wannabe* as a fixed slot are given in (68) and (69): in (68), *wannabe* follows a personal name (e.g. *Madonna wannabe*), whereas in (69) it precedes an adjective.

In these schemas, the double arrow represents the symbolic relation between form and meaning, the small caps subscripts denote part of speech, subscripts  $i$  and  $j$  are lexical indices and  $\{-, \_ \}$  denotes that the collocation can be written either as a single word, with a hyphen, or with a space in between (see Section 5.2.2).

(68)  $[[[x]_{\text{PNi}}\{-, \_ \}][wannabe]]_j \leftrightarrow [\text{aspiring to be like SEM}_i]_j$

(69)  $[[[wannabe]\{-, \_ \}][x]_{\text{ADjIj}}]_j \leftrightarrow [\text{approximating SEM}_i]_j$

Schemas are output-based and flexible, on the usage-based assumption that linguistic knowledge is shaped by usage and keeps on changing across the life-span in what Kemmer & Barlow (2000: ix) have termed a ‘feedback loop’. Schemas are furthermore assumed to be organized in a single network (Bybee 2010; Diessel 2019), the construct-i-con (cf. Jurafsky 1992: 8; Goldberg 2019: 36). On this view, there is no distinction between word formation rules and lexical items, because the construct-i-con contains both morphological schemas and specific instantiations. This implies complex and multi-dimensional networks with paradigmatic links between constructions (of varying complexity and schematicity), by virtue of formal and / or semantic similarities. In the case of *wannabe*, not only are the schemas in (68) and (69) linked to each other (because they share a fixed slot), but also to other constructions involving *wannabe*, or to other approximative constructions, to other constructions involving personal names, and so on.

A network approach offers a possible explanation for attested differences between English and the recipient languages. Univerbation of *want to be* took place in English, and *wannabe* can still be used as a verbal phrase (see Section 2.2.). In the other languages, *wannabe* was borrowed both as a noun and in specific collocations, but not as a verbal phrase (for instance, Dutch *?ik wannabe (een ster)* ‘I wannabe (a star)’ is not found in the nlTenTen20 corpus and sounds decidedly



odd). As a result, the place of *wannabe* in the construct-i-con and the way it is linked to other constructions is different in English. This may account for the observation that we find more collocations with human reference in English than in a language like Danish, where *wannabe* is more often found in combination with non-human collocations than in English. Interestingly though, we do not see network expansion (reflected by high type frequency) overall, but rather the emergence of specific clusters, e.g. food items in Danish, which suggest particular ‘pockets of productivity’ (Cappelle 2014).

The different network configurations for English as opposed to the recipient languages may also account for the differences in frequency distributions. We noted earlier that the high proportion of hapaxes reflects a schema with a high potential to expand, to include both new construction types and collocations with native nouns. As the recipient languages are not ‘hampered’ by the link to the VP *want to be*, as explained above, *wannabe* can perhaps more easily be combined with non-human nouns, where the implication ‘to aspire to (but probably in vain)’ is no longer possible. With inanimates, *wannabe* has acquired a more general meaning of approximation / fakeness but, because these constructions are still linked to *wannabe* constructions with human collocates, the latter cluster is also affected by the semantic shift in the former. As mentioned in Section 2.1, approximation can, in specific contexts, be interpreted as depreciation (Masini, Norde and Van Goethem 2023: 11). We find this with *wannabe* as well – see the two possible interpretations of *wannabe actress* in examples (59–60) above, or (70), where Stockholm is not so much approximating the concept of a capital city (it is the capital of Sweden after all), but it is dismissed as a bad specimen, much less worthy than Göteborg (Sweden’s second largest city). The depreciative sense can also be due to an association with the free noun *wannabe* as a ‘phoney, disingenuous person’ (see example (1)), as appears to be the case in example (71). This example, where *wannabe* is purely pejorative, is found in the context about Vietnam veterans inflating their heroism during the war.

- (70) [Danish] *Stockholm er en **wannabe-hovedstad**. I følge Tomas er det selvfølgelig Göteborg, der er den rigtige hovedstad i vores naboland.* ‘Stockholm is a **wannabe capital**. According to Tomas, Göteborg is, naturally, the true capital of our neighbouring country’
- (71) [English] “Who’d have thought that thirty years after the war all the records would be available to the public” Exactly, who ever thought that would happen? Certainly not all the **wannabe arseholes** who’d been getting away with their crap for that many years.

Apart from differences between English on the one hand and the recipient languages on the other, we also found differences within the group of recipient languages. In all languages, we saw both matter borrowing (e.g. *wannabe bad boy* found in the French sample) and pattern borrowing, where *wannabe* collocates with native words (e.g. French “*wannabe*”-*sauveurs de la terre* ‘wannabe saviours of the earth’). When there is more matter borrowing, we may expect stronger similarities to English in terms of word order, semantic category of the collocation and / or part of speech of the collocation.<sup>11</sup> At the same time, typological factors may be at play, particularly word order. Compounds in Germanic languages are generally head-final, which is why differences in word order are non-significant between English and Danish or Dutch. If, however, we find a similar non-significant result for French, which is head-initial, this implies stronger links to collocations that are matter borrowing, and hence (?) a lower degree of creativity in French.

Turning now to our model of F1-, F2- and E-creativity mentioned in Section 1, we review the most common English *wannabe* constructions in Table 4.

**Table 4:** English *wannabe* constructions and creativity.

Construction type			Creativity	
a	[[[x] <sub>PNi</sub> {-, <sub>u</sub> }][ <i>wannabe</i> ]] <sub>j</sub> <i>Hitler-wannabe</i>	↔	[aspiring to be like SEM <sub>i</sub> ] <sub>j</sub>	F1
b	[[[ <i>wannabe</i> ]{-, <sub>u</sub> }][x] <sub>PNi</sub> ]] <sub>j</sub> <i>wannabe Napoleon</i>	↔	[aspiring to be like SEM <sub>i</sub> ] <sub>j</sub>	F1
c	[[[ <i>wannabe</i> ]{-, <sub>u</sub> }][x] <sub>N{human}</sub> ]] <sub>j</sub> <i>wannabe gangsta</i>	↔	[aspiring to be like SEM <sub>i</sub> ] <sub>j</sub>	F1
d	[[[ <i>wannabe</i> ]{-, <sub>u</sub> }][x] <sub>N{inanimate}</sub> ]] <sub>j</sub> <i>wannabe road</i>	↔	[approximating SEM <sub>i</sub> ] <sub>j</sub>	F2
e	[[[ <i>wannabe</i> ]{-, <sub>u</sub> }][x] <sub>Ni</sub> ]] <sub>j</sub> <i>wannabe gangsta / road</i>	↔	[approximating SEM <sub>i</sub> ] <sub>j</sub>	F1

<sup>11</sup> We did not test differences in matter borrowing statistically, because there is no strict demarcation between matter borrowing and native words. Many loans, e.g. *hiphop* or *trendy* in Danish, are not necessarily perceived by speakers as foreign matter, which makes annotation for etymological origin of the collocate extremely difficult. Therefore, we treat matter borrowing as tendencies rather than hard figures.

Table 4 (continued)

	Construction type		Creativity
f	[[[ <i>wannabe</i> ] <sub>ADJ</sub> {-,~ } <sub>SEM<sub>1</sub></sub> ] <sub>NI</sub> ] <sub>J</sub> <i>wannabe arsehole</i>	↔	[bad specimen of SEM <sub>1</sub> ] <sub>J</sub> F2 (> E)
g	[[[ <i>wannabe</i> ] <sub>ADJ</sub> {-,~ } <sub>SEM<sub>1</sub></sub> ] <sub>NI</sub> ] <sub>J</sub> <i>actual and wannabe plutocrats</i> <i>I don't think she's wannabe</i>	↔	[approximating SEM <sub>1</sub> ] <sub>J</sub> F2 (> E)

Starting with collocations with personal names (a and b), we argue that both are F1-creative, i.e. fully productive and conventionalized word formation schemas.<sup>12</sup> Our synchronic data suggest that the same is true for collocations with nouns denoting humans (c). When the semantic category of the noun is not HUMAN, however, as in (d), we argue that this is an F2-creative extension. Because of the phrasal origin of *wannabe* (from *want to be*), which requires a human subject, we reconstruct a HUMAN > INANIMATE pathway. While such inanimate collocates are initially F2-creative, they may eventually lead to the more general schema (e), where the constraint that the *wannabe* collocate is HUMAN has been lifted. However, there is no new schema, so we argue that (e), once conventionalized, continues the F1-creativity of (c). In other words, these are cases of creativity followed by routinization. A different kind of F2-extension is found in (f): here, the semantic change is not in the collocate but in *wannabe* itself: a *wannabe arsehole* is not someone aspiring to be an arsehole, but just a terrible specimen of the collocate. This too is an incremental change, because depreciation is sometimes implicit in approximation, but it does imply that *wannabe* is becoming polysemous and if this goes hand in hand with a category change from noun to adjective, we have a new, E-creative schema. Finally, constructions that are clearly adjectival, like those in (g) can also be seen as F2-extensions from (e), but here, too, one might argue that the category change from noun to adjective is E-creative, especially in predicative constructions where *wannabe* is used independently (i.e. not as part of a collocation).

In the other languages, we see similar patterns, but changes in frequencies suggest that some languages are further advanced than English. Danish, for example, has less HUMAN collocates, which can be interpreted to mean that (e) (without

<sup>12</sup> Because (a) predates (b) in the OED and has higher type frequency, (b) may originally have been F2-creative, but this requires a further diachronic study.

the HUMAN constraint) is more strongly entrenched in Danish than in English. Since we also find more examples of adjectival usage, we may similarly argue that (f) and (g) are more entrenched and hence further down the E-creative road.

This view on creativity poses two challenges. The first one is that change is incremental, so that the line between F- and E-creativity is not easy to draw (Ungerer and Hartmann 2023: 44). The second challenge concerns the relation between creativity and extravagance (see for instance the papers in Eitelmann and Haumann 2022a). Where Eitelmann and Haumann (2022b: 5) see E-creativity as coming close to extravagance, Hoffmann (Hoffmann 2020) argues that some cases of E-creativity, e.g. the development of English *to be going to* as a future auxiliary, are not deliberate.

## 7 Conclusions and Outlook

The six languages studied in this paper show substantial variation in *wannabe* constructions at several levels (formal properties, semantic properties and productivity). Overall, *wannabe* in both English and recipient languages is used either as an independent noun or adjective or in collocations, where it tends to acquire an approximative (often derogatory) meaning. However, whereas in English *wannabe* collocates almost exclusively with nominal collocates (nouns, NPs, proper names), in the recipient languages we found expansion to unexpected categories (adjectives, verbs, and pronouns). On the semantic level, English differs from the recipient languages as well: in the latter we find more examples of inanimate collocates, presumably because speakers of languages other than English do not associate *wannabe* with the verbal phrase (*want to be*) it derives from.

The large number of hapaxes in our data suggests that *wannabe* is very productive while at the same time we see new schemata emerging. This is evidenced by tiny mismatches in semantics of either the collocate or *wannabe* itself, but we also found category shifts from noun to adjective. These are all tendencies, and stronger in some languages than in others, but we do argue that they reflect different kinds of creativity: mismatches are (initially) F2-creative, but they may result in either F1-creativity (generalization of an existing schema), or E-creativity (the emergence of a new schema).

Naturally, changes from F2 to F1 or E are gradual and we base our analysis on synchronic data. More data to test our three-way model of creativity is needed, not least diachronic corpus studies. Another interesting angle for future research is intralinguistic competition with other (morphological) markers of approxima-

tion. In any event, emerging phenomena like *wannabe* are especially interesting to be “photographed” to unveil the possible paths of variation in both creativity and routinization processes.

## References

The data that support the findings of this study are openly available in AMS Acta, the institutional open access repository of the Alma Mater Studiorum – University of Bologna, at <https://doi.org/10.6092/unibo/amsacta/8063>.

- Ainiala, Terhi & Jan-Ola Östman. 2017. Introduction: Socio-onomastics and pragmatics. In Terhi Ainiala & Jan-Ola Östman (eds.), *Socio-onomastics: The pragmatics of names*, 1–18. Amsterdam, Philadelphia: John Benjamins. <https://doi.org/10.1075/pbns.275>.
- Baayen, R. Harald. 2009. Corpus linguistics in morphology: Morphological productivity. In Anke Lüdeling & Merja Kytö (eds.), *Corpus linguistics. An international handbook*, 899–919. Berlin & New York: De Gruyter Mouton.
- Bauer, Laurie. 2019. Compounds. In Bas Aarts, Jill Bowie & Gergana Popova (eds.), *The Oxford handbook of English grammar*, 262–280. Oxford: Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780198755104.013.3>.
- Bergs, Alexander. 2018. Learn the rules like a pro, so you can break them like an artist (Picasso): Linguistic aberrancy from a constructional perspective. *Zeitschrift für Anglistik und Amerikanistik* 66(3). 277–293. <https://doi.org/10.1515/zaa-2018-0025>.
- Booij, Geert. 2010. *Construction Morphology*. Oxford: Oxford University Press.
- Booij, Geert. 2016. Construction Morphology. In Andrew Hippisley & Greg Stump (eds.), *The Cambridge handbook of morphology*, 424–448. Cambridge: Cambridge University Press.
- Booij, Geert. 2018. The construction of words: Introduction and overview. In Geert Booij (ed.), *The construction of words: Advances in Construction Morphology*, 3–16. Dordrecht: Springer International Publishing. [https://doi.org/10.1007/978-3-319-74394-3\\_1](https://doi.org/10.1007/978-3-319-74394-3_1).
- Bybee, Joan. 2010. *Language, usage and cognition*. Cambridge: Cambridge University Press.
- Cabin, Robert J. & Randall J. Mitchell. 2000. To Bonferroni or not to Bonferroni: When and how are the questions. *Bulletin of the Ecological Society of America* 81(3). 246–248.
- Cappelle, Bert. 2014. Conventional combinations in pockets of productivity: English resultatives and Dutch ditransitives expressing excess. In Ronny Boogaart, Timothy Coleman & Gijsbert Rutten (eds.), *Extending the scope of Construction Grammar*, 251–282. Berlin & New York: De Gruyter Mouton. <https://doi.org/10.1515/9783110366273.251>.
- Cappelle, Bert, Robert Dauts & Stefan Hartmann. 2023. The English privative prefixes *near-*, *pseudo-* and *quasi-*: Approximation and ‘disproximation.’ *Zeitschrift für Wortbildung / Journal of Word Formation* 7(1). 52–75.
- Diessel, Holger. 2019. *The grammar network: How linguistic structure is shaped by language use*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/9781108671040>.
- Dressler, Wolfgang U. & Lavinia Merlini Barbaresi. 1994. *Morphopragmatics: Diminutives and intensifiers in Italian, German and other languages*. Berlin & New York: De Gruyter Mouton.

- Eitelmann, Matthias, Kari E. Haugland & Dagmar Haumann. 2020. From *engl-isc* to *whatever-ish*: A corpus-based investigation of *-ish* derivation in the history of English. *English Language & Linguistics* 24(4). 801–831. <https://doi.org/10.1017/S1360674319000340>.
- Eitelmann, Matthias & Dagmar Haumann. 2022a. *Extravagant morphology: Studies in rule-bending, pattern-extending and theory-challenging morphology*. Amsterdam & Philadelphia: John Benjamins.
- Eitelmann, Matthias & Dagmar Haumann. 2022b. Introduction: Extravagance in morphology. In Matthias Eitelmann & Dagmar Haumann (eds.), *Extravagant morphology: Studies in rule-bending, pattern-extending and theory-challenging morphology*, 1–18. Amsterdam, Philadelphia: John Benjamins.
- Eitelmann, Matthias & Dagmar Haumann. 2023. *Getting close-ish*: A corpus-based exploration of *-ish* as a marker of approximation and vagueness: *Zeitschrift für Wortbildung / Journal of Word Formation* 7(1). 76–100. <https://doi.org/10.21248/zwjw.2023.1.91>.
- Gardani, Francesco. 2020. Borrowing matter and pattern in morphology. An overview. *Morphology* 30(4). 263–282. <https://doi.org/10.1007/s11525-020-09371-5>.
- Goldberg, Adele E. 2019. *Explain me this: Creativity, competition, and the partial productivity of constructions*. Princeton: Princeton University Press. <https://doi.org/10.2307/j.ctvc772nn>.
- Grandi, Nicola. 2017. I diminutivi come marche di attenuazione e indeterminatezza. In Oana-Dana Balaș, Adriana Ciama, Mihai Enăchescu & Anamaria Gebăilă (eds.), *L'expression de l'imprécision dans les langues romanes*, 162–175. Bucharest: Ars docendi – Universitatea din București.
- Grandi, Nicola & Livia Körtvélyessy. 2015. Introduction: Why evaluative morphology? In Nicola Grandi & Livia Körtvélyessy (eds.), *The Edinburgh Handbook of Evaluative Morphology*, 3–20. Edinburgh: Edinburgh University Press.
- Hoffmann, Thomas. 2020. Speakers are creative, within limits – a response to Peter Uhrig. *Cognitive Semiotics* 13(1). 1–7. <https://doi.org/10.1515/cogsem-2020-2028>.
- Hüning, Matthias, Ulrike Vogl, Ton Van Der Wouden & Arie Verhagen (eds.). 2006. *Nederlands tussen Duits en Engels. Handelingen van de workshop op 30 september en 1 oktober 2005 aan de Freie Universität Berlin*. Leiden.
- Jurafsky, Daniel. 1992. *An on-line computational model of human sentence interpretation: A theory of the representation and use of linguistic knowledge*. University of California, Berkeley.
- Kemmer, Suzanne & Michael Barlow. 2000. Introduction: A usage-based conception of language. In Michael Barlow & Suzanne Kemmer (eds.), *Usage-based models of language*, vii–xxiii. Stanford: CSLI Publications.
- Kempf, Luise & Matthias Eitelmann. 2018. Von “diutisk” zu “dynamisch”, von “englisc” zu “anything-ish”. “-is(c)h” kontrastiv diachron. *Zeitschrift für Wortbildung / Journal of Word Formation* 2(1). 93–134. <https://doi.org/10.3726/zwjw.2018.01.04>.
- Kilgarrieff, Adam, Vít Baisa, Jan Bušta, Miloš Jakubiček, Vojtěch Kovář, Jan Michelfeit, Pavel Rychlý & Vít Suchomel. 2014. The Sketch Engine: Ten years on. *Lexicography* 1(1). 7–36. <https://doi.org/10.1007/s40607-014-0009-9>.
- König, Ekkehard & Volker Gast. 2018. *Understanding English-German contrasts*. 4th edn. Berlin: Erich Schmidt Verlag.
- Kosse, Maureen. 2021. “Don’t make me pull a Britney”: Onomastics and genericness in the PULL A [PROPER NAME] construction. *Colorado Research in Linguistics*. <https://doi.org/10.33011/cril.v25i.1133>.
- Lamiroy, Béatrice. 2011. Degrés de grammaticalisation à travers les langues de même famille. *Mémoires de la Société de linguistique de Paris* 19. 167–192.

- Masini, Francesca & Jenny Audring. 2019. Construction Morphology. In Jenny Audring & Francesca Masini (eds.), *The Oxford handbook of morphological theory*, 365–389. Oxford: Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199668984.013.25>.
- Masini, Francesca & Maria Silvia Micheli. 2020. The morphological expression of approximation: The emerging *simil-* construction in Italian. *Word Structure* 13(3). 371–402. <https://doi.org/10.3366/word.2020.0176>.
- Masini, Francesca, Muriel Norde & Kristel Van Goethem. 2023. Approximation in morphology: A state of the art. *Zeitschrift für Wortbildung / Journal of Word Formation* 7(1). 1–26. <https://doi.org/10.21248/zwjw.2023.1.100>.
- Micheli, Maria Silvia. 2023. The emergence of approximative values in the Italian prefixes *semi-* and *para-*. *Zeitschrift für Wortbildung / Journal of Word Formation* 7(1). 156–176. <https://doi.org/10.21248/zwjw.2023.1.87>.
- Norde, Muriel & Graeme Trousdale. 2023. Issues in Diachronic Construction Morphology. *Constructions and Frames* 15(2). 145–159. <https://doi.org/10.1075/cf.00070.nor>.
- Norde, Muriel & Graeme Trousdale. 2024. Creativity, paradigms and morphological constructions. Evidence from Dutch pseudoparticiples. *Linguistics*. <https://doi.org/10.1515/ling-2023-0194>.
- Oltra-Massuet, Isabel. 2017. Towards a morphosyntactic analysis of *-ish*. *Word Structure* 10(1). 54–78. <https://doi.org/10.3366/word.2017.0100>.
- Sampson, Geoffrey. 2016. Two ideas of creativity. In Martin Hinton (ed.), *Evidence, experiment and argument in linguistics and philosophy of language*, 15–26. Berlin: Peter Lang.
- Uhrig, Peter. 2020. Creative intentions – The fine line between ‘creative’ and ‘wrong.’ *Cognitive Semiotics* 13(1). 1–19. <https://doi.org/10.1515/cogsem-2020-2027>.
- Ungerer, Tobias & Stefan Hartmann. 2023. *Constructionist approaches: Past, present, future*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/9781009308717>.
- Van Goethem, Kristel & Muriel Norde. 2020. Extravagant “fake” morphemes in Dutch. Morphological productivity, semantic profiles and categorical flexibility. *Corpus Linguistics and Linguistic Theory* 16(3). 425–458. <https://doi.org/10.1515/cllt-2020-0024>.
- Van Goethem, Kristel, Muriel Norde & Francesca Masini. 2025. The fate of ‘pseudo-’ words. A contrastive corpus-based analysis. *Languages in Contrast* 25(1). 23–50. <https://doi.org/10.1075/lic.22003.van>.
- Van Haeringen, Coenraad Bernardus. 1956. *Nederlands tussen Duits en Engels*. Den Haag: Servire.
- Vassiliadou, Hélène, Francine Gerhard-Krait, Georgia Fotiadou & Marie Lammert. 2023. Pseudo(-) in French and Greek: Categorization and approximation. *Zeitschrift für Wortbildung / Journal of Word Formation* 7(1). 234–262.
- Zeldes, Amir. 2013. *Productivity in argument selection: From morphology to syntax*. Berlin & New York: De Gruyter Mouton. <https://doi.org/10.1515/9783110303919>.

