

Online Appendices for “The rise of colligations: English *can’t stand* and German *nicht ausstehen können*”

Olav Hackstein and Ryan Sandell

International Journal of Corpus Linguistics

Appendix 1. Previous research on the diachronic evolution of synchronic collocations and colligations

Building upon Firth (1964) and Manning and Schütze (1999: 151–189), Gries and Stefanowitsch (2003) have refined the study of collocations through the application of quantitative corpus linguistic methods, including collocation analysis. Collocation analysis permits the quantification of the extent to which the usage of a particular lexeme is bound up with the usage of a particular construction. Thus, one can precisely assess whether a lexeme depends upon or triggers the employ of a given construction, which potentially allows for the discovery of diachronic archaisms through the synchronic quantitative assessment of collocation dependencies. Nonetheless, many insights following from quantitative corpus linguistics have not yet been integrated in the synchronic and diachronic analysis of multi-word phrases; for instance, recent treatments of German phraseology in Palm (1995), Burger (2007a, 2007b), and Donalies (2009) do not attempt to apply any quantitative approaches in their analyses.

Appendix 2. Comparative reconstruction of E. *stand* “tolerate, endure”, G. *ausstehen* “bear, tolerate” and persistent parameters

Inherited and persistent parameters of E. *stand* “tolerate, endure”, G. *ausstehen* “bear, tolerate” can be identified in three domains: a.) lexical semantics, b) the verb’s transitivity through preverbalization, and c) its stem formation.

a.) The lexical semantics of PIE **steh₂-* underlying E. *stand*, G. *ausstehen*

The lexical semantics of E. *stand* “tolerate, endure”, G. *ausstehen* “bear, tolerate”, as well as the verb’s polysemy between a verb of position and a verb of physical or cognitive endurance (by metaphorical extension) is found in Germanic, Slavic, Latin, Greek, and Vedic Sanskrit (see LIV² s.v. **steh₂-*), which warrants the projection of this polysemy with intransitive “endure” back into the Proto-Indo-European phase.

E. *can’t stand sb./sth.* contains the simplex verb *stand* in the metaphorical sense of “be strong and endure,” which continues the semantics of the PIE root **steh₂-* “stand (upright) firmly”. A similar metaphorical extension, producing the sense “endure” is also found in other, earlier attested Indo-European languages, for instance:

- i) Latin **verb of position** *stare* “to stand, be in a standing position, stand up (of people, animals)” (Glare’s, 2012, *Oxford Latin Dictionary*, henceforth “OLD”, s.v. *stō*, 1), “to stand up (in or for combat)” (OLD, s.v. *stō*, 1 d 2) → **verb of endurance** “to remain stable, last (of a city institution etc.)” (OLD, s.v. *stō*, 16), “to continue to exist, endure” (OLD, s.v. *stō*, 17).
- ii) Greek **verb of position** ἵστημι [hístē:mi] causative “make to stand, set up” (Liddell et al., 1925–1940, henceforth “LSJ”, 841, s. v. ἵστημι, sub A causal), anticausative ἵσταμαι [hístamai] (with aorists ἔστην [éste:n], ἔστηκα [héstē:ka]) “to be set up, erected” (LSJ, 841, s. v. ἵστημι, sub B II 4) → **verb of endurance** ἵσταμαι [hístamai] “to stand still [and endure]” (LSJ, 841, s. v. ἵστημι B II).

b.) Transitivity of **steh₂-* through optional preverbal reinforcement and reinforcement with PIE **ud/ūd/ut-s*

Possibly already in Proto-Indo-European, and certainly in Proto-Germanic, **steh₂-* could be optionally reinforced with a local preverb: PIE **ud* (Ved. *úd* “up, upon”), **ūd* “on high” (with monosyllabic lengthening in PGmc. **ūt* “out”), and directional *ut-s* “upward” (in OCS *vŭz*, Goth. *us*, ModG. *er-*) can be found in together with **steh₂-*, thereby yielding a collocation **ud/ūd/ut-s steh₂-* “stand (up to sb./sth.) firmly, endure”. Examples of the local and directional particles with **steh₂-* follow below.

- i) **Local PIE **ud* (Ved. *úd*), **ūd* “on high”:** Vedic *úd sthā-* “stand up [against an obstacle]”, hence “endure, withstand”: RV 4.4.4a: *úd agne tiṣṭha* “stand up, Agni [against your enemies]”, RV 10.53.8b: *út tiṣṭhata* “stand up (endure)!””, RV 8.76.10: *úd tiṣṭhan ójasá sahá* “standing up [enduring] with your might”. This Vedic usage of *ud sthā-* entirely parallels the sense of “endure” attested throughout the modern Germanic languages.
- ii) **Directional PIE **ut-s* > **uss* > **us*.** Examples of directional **ut-s* include OCS *vŭs-stati* (Kurz’s, 1958–, *Lexicon linguae palaeoslovenicae*; “LLP”, 1.339 s.v. *vŭs-stati* 2) *na kogo* “to rise up against sb.” [sich gegen jem. erheben]), Russ. *vosstat’* “to rise against, withstand”, Goth. *us-standan* “rise, be resurrected (with endurance)”, and OHG *ir-standan* “rise, be resurrected” (cf. further 3.2.3.1 below). The function of the morpheme *-s* in PIE **ut-s* “upward” is not certain (cf. Schneider 2011: 189f., Dunkel 2014: 823f., 825, Opfermann 2016: 250–2, Kroonen 2010: 370), but it is likely directional (cf. Hackstein 2002: 109 fn. 12), or less likely ablatival “from up(right)”; it is worth noting that PIE **(h₁)en-s* means “inwards, into”, not “from inside”.

Repetitive reinforcement caused PGmc. **uz* to become **ūt-uz-*, which is continued by Gothic *ut-us-*, and OHG *ir-stantan* → *ūz-ir-stantan* (Schützeichel, 2012: 310); cf. ModG *er-koren* “choose”, *aus-er-koren* “chosen”, ModG. *etwas erstehen* with semantic shift “to reach sth. by endurance”. The directional function of **ut-s* as continued in G. *er-*, against the local function of **ūt* as continued in G. *aus-*, also accounts for the fact that the bound prefix G. *er-* occurs as a telic-transitivizing morpheme in German and is more strongly transitivizing than G. *aus-*. Contrast a) G. itr. *steigen* “climb” vs. tr. *er-steigen* “to reach/accomplish sth. by climbing” with b) G. intr. *steigen* “climb” vs. intr. *aus-steigen* “to get out, get off”.

Given the intransitive meaning of the base lexeme PIE **steh₂-*, Gmc. **stō-* “stand (upright) firmly”, the transitive use of E. *stand*, Du. *uitstaan*, G. *ausstehen* as “endure sb./sth.” is anomalous and hints at a persistent archaism in accordance with Meillet’s (1937) heuristic principle that persistent linguistic structures tend to appear synchronically as anomalies. In each case, however, the proof must be provided by comparative reconstruction. The synchronically anomalous transitivity of **steh₂-* in E. *stand*, Du. *uitstaan*, G. *ausstehen* may be attributed to two factors: addition of preverbs **ud* or **ut-s*, and the specific stem formation built in Proto-Germanic (see c.) below).

Table A1 illustrates some instances in both Modern and Old Indo-European languages in which preverbatation results in transitivity. Although **ud* and **ut-s* may not have always had a strongly transitivity effect (cf. Schneider, 2011: 185, and Kulikov, 2012: 734 sub 25 on Vedic Sanskrit *úd*), the West Germanic languages have developed a common optionally transitivity function: such examples include German *aussitzen* “sit out (smth.)” (< intrans. *sitzen* “sit”), *ausspielen* “finish playing; play somebody off against somebody” (< intrans. *spielen* “play”), or English *outrank* (< intrans. *rank*).

Table A1. Transitivity effect of preverbs on “go”

Language	V _{itr}	→ V _{tr}	Gloss
German	<i>gehen</i>	<i>über-gehen</i>	“transgress, ignore”
Russian	<i>idti</i>	<i>pere-jti</i>	“cross over”
Latin	<i>gredi</i>	<i>ag-gredi</i>	“go towards, attack”
Hittite	<i>pai-</i>	<i>istarna arha pai-</i>	“go through” (Cotticelli-Kurras, 2007: 13)

c.) Morphological transitivity of **steh₂-* by its stem formation (nasal infix)

Given that the Present-day English transitive usage of *stand* does not clearly continue a virtual **stand out* “endure”, another mechanism other than preverbatation might be responsible for the transitivity of **steh₂-* in Germanic as well. The Proto-Germanic stem formation **standa-*, a nasal-infix present with dental enlargement (see Prokosch, 1939: 67, 157; Lühr, 2000: 94–95; Rix et al.’s, 2001, *Lexikon der indogermanischen Verben*, henceforth LIV² 591 fn. 1; Mottausch, 2015: 99, and Ringe, 2017: 277) is a likely culprit. PGmc. **standa-*, as a nasal present, continues a PIE causative function “to place”, which, in middle voice reads as “to place oneself, step, stand”. On the transitivity function of the nasal present in PIE, see Meiser (1993), Villanueva-Svensson

(2011), Covini (2016/7: 290–328 [326–8]), Zasada (2020). Scheungraber’s (2014: 69–75) alternative proposal to derive the stem of Germanic *stand* from an *nt*-participle is less persuasive because it leaves the verb’s lexical semantics unexplained; see also Mottausch (2015: 97).

In the active voice, PGmc. **standa-* and **ūt/ūs standa-* meant “to cause sb/sth to stand upright; to put up, erect”, as well as metaphorically “to be able to sustain, to bear, endure”. For similar metaphorical extensions using other roots, compare PIE **telh₂-* “to lift”, “to be able to lift and endure”, underlying Hom. Gk. οὐκ ἔτλη [o:k étlɛ:] “he couldn’t bear”, or PIE **b^her-* “carry, bear”, underlying Latin *non ferō* “I can’t stand (something)”. In the middle voice, PGmc. **stand-* and **ūt/ūs stand-* meant “to cause oneself to stand upright (towards)”, “to stand up (to)”, and metaphorically “to endure”; in addition, the phrase began to be construed with an accusative object denoting the goal, with the meaning “to stand (up) to somebody or something”. The goal adjunct came to be reanalyzed subsequently as a direct object, increasing the valency of **steh₂-* to that of a two-place predicate. For a Germanic parallel for the transition of a motion verb governing the accusative of goal and its semantic development to a psych verb, see Appendix 3 below on G. *leiden*.

Old Norse still semantically mirrors the middle forms in *standa* “stand, be steadfast”, *standask*, *stózk* “to let oneself stand out against; withstand, endure”; likewise, Modern Icelandic *standast próf* means “to pass an exam”, and *standa sig* “to maintain one’s hold”. The same mechanism, namely, the shift from a motion verb to an experiencer verb, also accounts for the transitivity of English *under-stand*, OE *understandan*, Frisian *understonda*, Middle Dutch *understēn/understān* from PIE **(h₁)ntér steh₂-*, PGmc. middle **under standa-* “cause oneself to stand in between/amidst something, and be able to perceive and comprehend it”. Note here that *under* reflects not PIE **nd^hér* “under, beneath”, but **(h₁)ntér* “in between”, which merged with PIE **nd^hér* in Germanic; see Harm (2003: 115).

Appendix 3. E. *stand* / G. *ausstehen* and their lexical competitors: history and etymology

The sketch in Appendix 2 of the linguistic prehistory behind E. *stand* and G. *ausstehen* strongly suggests that the functions of these verbs found in the present-day collocations *can't stand* and *nicht ausstehen können* reach back into at least into Proto-West Germanic, but likely even into Proto-Indo-European. A closer look at the lexical competitors of *stand* and *ausstehen* in the sense “endure”, E. *tolerate*, *endure*, G. *leiden*, *ertragen*, *aushalten*, *mögen*, whose colligational relationships to the CAN NOT VERB_{INF} X₀ OBJ construction are explored quantitatively in Section 5 of the main text, largely reveals the opposite to be true: their usages as transitive verbs meaning “endure” are comparatively innovative, or they constitute recent lexical borrowings; see the following etymologies.

a) English

stand

Intransitive *stand* is, as shown under Appendix 2, reconstructable for Proto-Indo-European, and well-attested from Old English onward (see Bosworth-Toller, 1921, s.v. *standan*), though transitive usages are first clearly found in the 14th century (cf. MED s.v. *stōnden*). Note that English originally possessed a verb *astandan* (see Bosworth-Toller s.v. *astandan*, MED s.v. *astōnden*, OED s.v. *astand*), cognate with OHG *irstandan* (cf. above), which may have fallen together formally with *standan* via apheresis of the prefix *a-* (see OED s.v. *a-*, prefix₁). Given OHG *irstantan* and Gothic *usstandan*, a Proto-Germanic **uz-standa-* likely existed, and perhaps could have been employed in the meaning “endure, withstand” besides the more robustly attested sense “arise, stand up”.

bear

Alongside E. *stand* “endure, tolerate” we find synonymous E. *bear* “endure, tolerate” which, as a lexeme is of equal linguistic age and can be projected back into Proto-Indo-European; cf. Lat. *ferre* “endure, tolerate”, Greek φέρω “endure, tolerate” (LSJ 1923 s.v. III, e.g. Homer *Odyssey* 18.135), and Vedic Sanskrit *bhar-* “endure, tolerate” (Grassmann, 1872 [1976]: 956 No. 14 “zu erfahren haben”).

In accordance with its linguistic age, E. *bear* “endure, tolerate” exhibits many of the same syntagmatic constraints as E. *stand* “endure, tolerate”, see, e.g. the colligation of E. *bear* “endure, tolerate” with *can’t*. One cannot say **She bore to be neglected/ being neglected*, while *She can’t bear to be neglected/ being neglected* is grammatically acceptable (cf. Bolinger 1976:9). The quantitative collocation analysis under Section 5 likewise indicates that *bear*, like *stand*, is more attracted to and dependent upon the CAN NOT VERB_{INF} X₀ OBJ construction than the other verbs investigated. These two verbs have, however, specialized somewhat differently, in that *stand* exhibits a predilection for pronominal objects, while *bear* prefers infinitival complements.

tolerate, endure

E. *endure, tolerate* are loan verbs (see OED s.v. *endure* and *tolerate*), first attested in English in the late 14th and 16th centuries respectively. They thus represent the linguistically youngest layer in English, in contrast to the older, inherited verbs *stand* and *bear*. Consequently, they are largely free of derivational constraints.

b) German

ausstehen

G. *ausstehen* with stressed separable prefix *aus-* has an Indo-European pedigree. It reflects the collocation of the local prefix PIE **ud*, PGmc. **ūt* and PIE **steh*₂-. While in both OHG and MHG, *stān/stān* occurs with a separable prefix *ūz/u3* in a wide variety of meanings (from “jump up” to “rest”), the meaning “endure, tolerate” does not happen to be one of them. By contrast MHG *erstān* (< OHG *irstān*) is attested with the sense “endure, tolerate” (cf. 4.2 in the main text). Therefore, G. *ausstehen* likely replaced older MHG *erstān* (< OHG *irstān*) with unstressed inseparable prefix OHG *ir-* from PGmc. **us* and PIE **ut-s*. The collocation of PIE **ut-s* with PIE **steh*₂- is attested in Germanic, Slavic, and Vedic (cf. Appendix 2).

leiden

G. *leiden* in *nicht leiden können* derives etymologically from the motion verb OHG *līdan*, Goth. *ga-leiþan* “come, go”, PGmc. **leiþe/a-* (PIE **leiþ*₂- “go”; see LIV²: 410). The underlying verb of motion could take an accusative of goal in the sense of “to go to/reach sb./sth.”, and metaphorically as “to incur something”, with a semantic shift from a motion verb governing the accusative of goal

to an experiencer and thus psych verb; cf. G. *leiden* in the idioms *Schaden leiden* “incur and suffer damage”, and *jemanden nicht leiden können* “not be able to tolerate sb.”. The semantics and the construction of G. *leiden* “endure” first arose in the High German of the 8th or 9th century. The verb is not as old as the forebears of G. *ausstehen*.

ertragen

NHG *ertragen* appears to be a comparatively recent coinage. It is unattested in OHG and MHG, and its first attestations stem from the early NHG period. See the DWB 3.1031, s.v. *ertragen*, sub 2 *sustinere, tolerare*.

aushalten

Like NHG *ertragen*, the first attestations of NHG *aushalten* in the sense of “withstand, tolerate” come from early NHG. See the DWB 1.879f., s.v. *aushalten, sustinere*, sub 4) mit acc. der person, and sub 5) mit acc. der sache: *den kampf, krieg, angrif, sturm, das spiel aushalten*.

mögen

The same applies to German *mögen* “like” and *nicht mögen* “dislike”, which belong to the semantically most recent layer of the modal *mögen*, arising first in NHG. The senses “like” and “dislike” presuppose the semantic shift of *mögen* “be able”, as attested for OHG *magan*, MHD *mügen* and preserved in NHG *vermögen* “be able”.

Appendix 4. The attested history of English *can't stand* and German *nicht ausstehen können*

This appendix accompanies Section 4 of the main paper and serves primarily to provide further examples of contextual usage of the verbal lexemes under discussion, as well as some basic statistics on frequencies of occurrence in different periods of English and German.

A4.1 English: (*can't*) *stand someone/something (out)*⁴

Old English *standan* as a **verb of position** is also illustrated in (A1); compare (4) in the main text.

(A1) ... þā **stōdan** him twegen weras bīg (*sic*) on hwītum hræglum. (*Ascension Day Homily*, 99–100; DOEC)

“... there stood before him two humble men in white clothing”

(A2) also attests to the semantics of **physical persistence or endurance**; cf. (5) in the main text.

(A2) ... and fela samod tugon, ac heō næs āstyrod, ac **stōd** swā swā munt. (*Ælfric's Lives of Saints*, 100; DOEC)

“... and many are drawing together, but she is not moved, but rather stood her ground (lit. mountain).”

(A3) represents another instance of OE *astandan* with the reading “endure”; cf. (5) in the main text.

(A3) forðæm is ðæm lāce swīðe geornlice to giemanne ðæt he swā strangne lācedom selle ðæm sēocan, swā he mæge ða mettrymnesse mid geflieman, & eft swā liðne swa se tȳdra lichoma mæge **astandan** (*Gregory's Pastoral Care* 61.455.28; DOE s.v. *astandan*)

“Because it is for that doctor very desirable to take care that he give such a strong medicine to the sick one, so that he can drive that illness away with it, and thus can the frail body pleasantly so endure”

Nevertheless, the lexical possibilities for the extended sense of mental or emotional endurance “tolerate, endure” in the Old English are mainly represented by a range of other verbs, including (*a-*)*beran* “bear” *a-drēogan* “perform, carry out; endure, suffer”, and *ar-æfnan* “suffer, endure, hold out”. That lexemes with the meaning “endure, tolerate” so often exhibit a prefix related to the preposition “out” suggests that the semantics of “out” could often compositionally construct the semantics of mental endurance. In particular, *aræfnan* seems to have been the default lexeme for expressing the meaning “endure, tolerate” in the Old English period, also forming part of an expression *aberan and aræfnan* “to bear and endure” (cf. DOE s.v. *aræfnan*); this lexeme is, however, evidently extinct by the Middle English period.

Examples (A4) and (A5) complement examples (7) and (8) in the main text, illustrating ME *stōnden*.

(A4) Fyrumbras was aggreued sare þat O[lyuer] hym **stod** so longe (c. 1380, *English Charlemagne Romances I: Sir Ferumbras*, 634; MED)

“Ferumbras was sorely aggrieved that Oliver (with)stood him so long”

(A5) Hij ne mowen þat assaut stonde, And fleizeþ a litel by þe stronde (c. 1400, *King Alexander* 3732; MED)

“he could not (with)stand that assault, and escaped a little by the beach”

Although the sense “withstand, resist, endure” is found for both *stōnden* and *astōnden* in Middle English, this usage is unattested in the EEBO corpus (1470–1700) of late Middle and Early Modern English until c. 1600. Transitive usages of *stand* in the EEBO between 1600 and 1700 predominantly attest to resistance or endurance of an explicitly physical kind (spellings here are normalized): *stand a blow*, *stand a/the fight*, *stand the test*, *stand a/the shock*, *stand the charge*, *stood the (enemy’s) armies*, *stood the assaults*, *stand a/the trial*, *standing the brunt (of the battle / of continual ceremony)*, *stand the force*, *stand the fury*. See also Jespersen (1927:345 §16.7,6). Example (A6) complements (9) in the main text.

(A6) ...deborah in person out-braving danger, and **standing the brunt of the battell**, against many thousands, living armed and awake... (1640, *The exemplary lives and memorable acts of nine the most worthy women in the world*; EEBO).

While the construction *stand out* OBJECT, appearing after c. 1600, is largely equivalent to *stand* OBJECT, the two differ somewhat in that is that the former is very frequently attested in the form *stand it out* “endure (something)” (2× in the 1580s, 315× (!) in the 17th century), while pronominal objects with simple *stand* are comparatively rare (see two clear instances in (A8) and (A9)) in this period. Examples (A7)–(A11) attest to the same pattern seen in examples (11) and (12) of the main text.

(A7) Hundreds he sent to Hell, and **none durst [dare] stand him** (a1616 *Shakespeare Henry VI, Pt. 1* (1623) i. i. 123).

(A8) haru: oh ware what loue is? ned hath found the scent; and if the connie chance to misse her Burrough, Shee’s Shee’s ouer-borne y fayth, she **can not stand it** (1616, *English-men for my money: or, a pleasant comedy called, A woman will haue her will*; EEBO).

(A9) pal: any thing: sink, ruin, perish: fate has not that frown, nor heaven and all its thunder has that bolt, but i **could stand ‘em** all for dear jacintha (1697, *The world in the moon, an opera*; EEBO)

(A10) i from those eyes for ever will remove, i **can not stand the sight** of hopeless love (1673, *The Empress of Morocco, a tragedy*; EEBO)

(A11) Jesus fled from the persecution; as he **did not stand it out**, so he **did not stand out against it** (1649 Bp. J. Taylor *Great Exemplar* I. vi. 105).

Given that the *stand* OBJECT and *stand out* OBJECT constructions in Early Modern English still exhibit a predilection for objects that allow the reading “withstand”, it seems plausible that these usages directly or effectively continue OE *astandan*. That *stand out* OBJECT should have this same particular preference, even though, as noted above, no OE **(a)standan ūt* or ME **(a)stōnden out(e)* seems to be attested, is peculiar, unless one presumes that such a particle verb did in fact exist prior to Early Modern English as a vernacular alternative to the *astandan* of the written language.

Alternatively, the formal merger of OE *standan* and *astandan* in *stand* provided an opening for the creation of the particle verb *stand out* as a potential replacement for *astandan*.

Once *stand* had developed the sense of a psych verb indicating **mental/emotional endurance**, the relevant sense became increasingly restricted to the context provided by the CAN NOT VERB_{INF} X₀ OBJ construction in the 18th and 19th centuries. For example, the COHA attests 242 instances of CAN NOT *stand* PRONOUN in the period 1800–1900, versus 201 instances of CAN *stand* PRONOUN without negation and certainly yet fewer instances of *stand* PRONOUN, without either negation or a preceding form of CAN. Examples (A12)-(A15) parallel examples (13) and (14) of the main text.

(A12) Till I am satisfied in these particulars, you and I must by no means meet; **I could not possibly stand it** (1750 Ld. Chesterfield *Let.* 19 Nov. (1932) (modernized text) IV. 1621).

(A13) **She could not** stand that Manager fellow. **I could not** stand him myself (1879 M. Oliphant *Within Precincts* II. Xx. 60).

(A14) Captain Blunt jumped up. “My mother **can’t** stand tobacco smoke” (1919 J. Conrad *Arrow of Gold* iv. li. 162).

(A15) Houses... many times **cannot** well **stand out** a long Lease (1676, H. Phillipps *Purchasers Pattern* 18)

For further details on the relevant usage of *stand*, see also OED s. v. *stand-* V. Transitive senses, **52.**, **58.**, and **59**. Compare (A16) and (A17) with (14) in the main text on the idiom *stand out* in British English; cf. also Swedish *Jag tar inte ut längre med henne* “I can’t stand it with her any longer”.

(A16) We managed to stand out against all attempts to close the company down (Hornby & Cowie, 1989:1250).

(A17) He is in good earnest, and will execute these threatnings upon them if **they will** obstinately stand it out with him (a1694 J. Tillotson *Serm.* (1742) III. xxxv. 17).

A4.2 German: From *(ir)stantan* to *nicht ausstehen können*⁸

(A18) shows OHG and *stantan* in its sense as a basic verb of position.

(A18) Andares tages abur **stuont** Iohannes inti fon sinen iungiron zuene inti giscouuota then heilant gangantan... (Tatian 16; RA)

“But on the second day John and his two disciples stood and saw the savior going by...”

To example (22) in the main text, it may be added that Pennsylvania German, as an isolated dialect of German, still maintains unrestricted usages of the verb *ausstehen*, i.e., Pennsylvania German *ausschteh*; see Beam (2004–2011) 1 A-152 for the examples in (A19) and (A20):

(A19) Sie will’s geduldich ausschteh.
“She wants to suffer through it patiently.”

(A20) Der Hi hot gemeht, des waer schlimm, en Mensch so ausschteh losse.
“He was of the opinion that it would be wrong to permit a human being to suffer so.”

Paralleling the restriction of *ausstehen* to use after modal verbs alongside (24) is (A21), while (A22) pairs with (25), indicating the further restriction to a negated construction with *können* specifically.

(A21) es ist aber einer, der dir deinen trotz wol **kann ausstehen** (16th c., Luther 6, 226b).
“...he is, however, one who can tolerate your contrariness well”

(A22) nein gnädger herr, sie **kann ihn nicht ausstehn**. (18/19th c., Tieck 3, 203)
“no, gracious lord, she cannot stand him.”

Appendix 5. Data collection for assessing colligational strength in the CAN NOT VERB_{INF} X₀ OBJECT_{ACC} construction

This appendix describes the query procedures used for obtaining the data for colligational analysis in Section 5.2 of the main text from the *Corpus of Contemporary American English* (COCA) and the subcorpora TAGGED-C and TAGGED-C2 of the *Deutsches Referenzkorpus* (DeReKo).

5.1 English

To obtain a relatively complete picture of the frequency of different types of elements in the X₀ OBJECT slot(s) of the transitive CAN NOT VERB construction, the following parts of speech were searched for in the COCA interface following the VERB slot for each collexeme (the relevant COCA query sequence is given in parentheses): pronoun (_p*), noun (_nn*), adjective followed by a noun (_j* _nn*), possessive pronoun followed by noun, with or without an intervening adjective (_app* [_j*] _nn*), verbal gerund (_v?g*), determiner or quantifier followed by a noun with or without an intervening adjective (_d* [_j*] _nn*), article with or without an intervening adjective (_at* [_j*] _nn*), or negation (_xx*). These possible X₀ OBJECT sequences were preceded by ca|can|could _xx* VERBAL-COLLEXEME (e.g. STAND, BEAR) in the search query. Thus, as an example, the complete query ca|can|could _xx* STAND _nn* returns all instances of the sequence *ca, can, or could*, followed by either *not* or *n't*, followed by any form of the lexeme *stand*, followed by any noun. Queries were run through both the List and Collocates functions of the COCA interface. The total frequencies of the respective verbal collexemes were obtained through queries of the form STAND_v, BEAR_v, etc. An approximate token frequency for all verbal constructions in the COCA was obtained through the query _v*; this search yielded a total token frequency of 100535787. Similarly, an approximate token frequency for the transitive CAN NOT VERB construction was obtained through queries with simply _v* in the slot of the verbal collexeme; this yielded a token frequency of 161174.

5.2 German

Since the German verbal collexemes examined here all obligatorily require a direct object, including particular types of nominal phrases was not necessary in the search term; throughout,

two different search patterns were employed to find instances of our colligation in main and subordinate clauses, respectively (VERB here is to be read as a verbal lemma, e.g. *ausstehen* or *machen*):

- Main Clause: ((&können %w0 (, oder ? oder .)) /+w5 "nicht") /+w1 &VERB /w0 MORPH(V INF)
- Subordinate Clause: "nicht" /+w1 (&VERB /w0 MORPH(V INF) %w0 (, oder ? oder .)) /+w1 &können

For a main clause, the search term finds all instances of the lemma *können* followed by *nicht* and a given verbal lemma in its infinitive, with up to five words separating *können* and *nicht*; punctuation (an orthographic indication of a clause boundary) cannot intervene between *können* and *nicht*. For a subordinate clause, the search term finds all instances of *nicht* immediately followed by a given verbal lemma in its infinitive, in turn immediately followed by a form of *können*; punctuation cannot intervene between the infinitive and *können*. To obtain a count of all instances of verbal forms in the two subcorpora, the simple search MORPH(V) was applied, yielding 261158216 tokens. Searches for the X₀ OBJECT_{ACC} NICHT VERB_{INF} KÖNNEN construction generally returned 362916 tokens, where the search term MORPH(V INF) was used without the specification of a lexeme.

Appendix 6: Productivity, colligational strength, and the age of constructions

It is a recognized fact that higher degrees of cohesion and greater fixity of an idiom (i.e. fewer possibilities to modify the elements of a multi-word construction) are indicative of comparatively older constructions, all things being (semantically and pragmatically) equal; cf. the well-researched correlation between chunking and morphological freezing (Bybee, 2006, 2007, 2010) and univerbation (Hackstein, 2012a), and the notion of *Festigkeit* “cohesion, fixity” in idioms according to Burger (2007a, 2007b). For similar results on Homeric Greek, see Bozzone (Forthcoming). Hence, a construction with a quantitatively lower relative productivity may be reasonably regarded as diachronically older. This does not mean that “old” constructions are necessarily “unproductive” in the sense that they cannot generate new types altogether, but merely that they may be expected to be comparatively *less* productive than other constructions that fulfill the same function. In the case at hand, we indeed find that higher colligational strength (on a variety of measures) correlates with lower productivity (as measured by Baayen’s \mathcal{P} ; see Baayen, 1989, 1992). This correlation suggests that particular collexemes ride similar diachronic waves, in both becoming more bound to particular constructions and exhausting the potential for the construction to generate new types.

The basic notion, as applied by Bozzone (Forthcoming), is that diachronically older constructions (and in particular, constructions that are “in decline” and are in the process of being replaced by functionally similar constructions), tend to lose, rather than gain, types, while a few established instances of the construction represent an increasing proportion of the tokens. This view then lends itself to the application of the corpus-based measures of morphological productivity developed by Baayen (1989, 1992; see also Sandell, 2015, in the context of historical linguistics), in particular the statistic \mathcal{P} (= HAPAX LEGOMENA / TOKENS), as a means of potentially comparing the relative age of functionally parallel, competing constructions.

Zeldes (2012) has already previously demonstrated that the same corpus-based measures of morphological productivity developed in Baayen (1989, 1992) can be appropriately applied to syntactic units. Although defining exactly what units count as distinct types or as tokens of the same type in syntactic units is not as straightforward as in morphology, Zeldes (2012: 106) comes to the conclusion that a distinct type for a syntactic construction should be defined as “a distinct realization which fills all open slots in a construction, where “distinct” and “construction” should

be understood in the CxG [Construction Grammar] sense, as entries in the mental lexicon.” As applied to the CAN NOT VERB_{INF} X₀ OBJECT_{ACC} construction, since three open slots occur (VERB_{INF}, X₀, and OBJECT_{ACC}), any distinct element in any one of those slots counts as a distinct type. Thus, in English, any combination of *bear* and *tolerate* in the VERB_{INF} slot, *that* and \emptyset in the X₀ slot, and *rabbit* or *rabbits* in the OBJECT_{ACC} slot would count as a distinct type.

For the case at hand, the productivity of specific collexemes in the VERB_{INF} can then be evaluated by “filling” the VERB_{INF} slot with *stand*, *bear*, etc. to generate more specific constructions, e.g. CAN NOT STAND X₀ OBJECT.¹³ In English, we assume for the present that the occurrence of *can not* versus *can’t* is governed by phonology and speech register or other sociolinguistic considerations, not morphology or syntax per se, and thus *can not stand rabbits* and *can’t stand rabbits* count as instances of the same type. Since this paper is principally concerned with the diachronic status of *can’t stand* and its lexical competitors, data on type frequency and hapax legomena was only gathered for *can’t stand*, *can’t bear*, and *can’t tolerate* (for English), and *nicht ausstehen können*, *nicht aushalten können*, *nicht leiden können*, and *nicht ertragen können* (for German).

From the same queries to the COCA corpus described in Section 5.2.1 of the main text, the number of *hapax legomena* for each of the collexemes *stand*, *bear*, and *tolerate* in the CAN NOT VERB_{INF} X₀ OBJECT_{ACC} construction was collected, which, together with the token frequencies of each colligation, allows for the calculation of Baayen’s \mathcal{P} ; these values, as well as the type frequency (V) of each colligation, are given in Table A2.

Table A2. Productivity (\mathcal{P}) of CAN NOT STAND/TOLERATE/BEAR.

	CAN NOT STAND X ₀ OBJECT _{ACC}	CAN NOT TOLERATE X ₀ OBJECT _{ACC}	CAN NOT BEAR X ₀ OBJECT _{ACC}
Types (V)	779	291	348
Tokens (N)	2266	340	818
<i>hapax legomena</i> (h1)	512	226	229
Baayen’s \mathcal{P} (h1/N)	0.2259	0.6647	0.2799

The proportion of *hapax legomena* to tokens for these three collexemes in the CAN NOT construction shows that the productivity of CAN NOT TOLERATE is considerably greater than CAN NOT STAND or BEAR. This result suggests, on the one hand, that the potential of CAN NOT TOLERATE is comparatively underexploited in Present-day English, and on the other, that CAN NOT STAND or BEAR may be substantially diachronically older than CAN NOT TOLERATE. Such a conclusion is perhaps unsurprising, given that *bear* and *stand* are inherited into English from PGmc. and PIE (see Section 3 of the main text), while *tolerate* is a loanword, first attested in English in the 16th century (OED s.v. *tolerate*). The somewhat lower productivity of CAN NOT STAND as compared to CAN NOT BEAR may in turn indicate that the tendency of STAND to be colligated with the CAN NOT construction is relatively older.

Furthermore, the statistic \mathcal{P} in the English data correlates strongly with the measures of colligational strength calculated in Table 8 in the main text. For instance, Pearson's R between \mathcal{P} in Table A2 and the respective values for MINIMUM SENSITIVITY is -0.814, and between the χ^2 likelihood ratio is -0.886. In other words, the stronger the colligational strength between a given collexeme and the CAN NOT VERB_{INF} X₀ OBJECT_{ACC} construction, the lower the productivity of the constructional instantiation with the collexeme. Insofar as the thesis of a relationship between smaller values of \mathcal{P} and greater linguistic age is valid, the conclusion may be drawn that English *stand* is older and more established in connection with the CAN NOT VERB_{INF} X₀ OBJECT as compared to *bear* and *tolerate*.

As far as the situation in German is concerned, Table A3 shows the same statistics given in Table A2, but for the collexemes *ausstehen*, *aushalten*, *leiden*, and *ertragen*. Due to the design of the search functions in COSMAS-II and the (in comparison to English) greater variety of syntactic structures, the types and hapax legomena for the German data had to be gathered by hand: each token of X₀ OBJECT_{ACC} instance was input into a spreadsheet, and then filtered for duplicate values (to produce a type count) or unique values (to produce a count number of hapax legomena).

Table A3. Productivity (\mathcal{P}) of NICHT AUSSTEHEN/AUSHALTEN/LEIDEN/ERTRAGEN KÖNNEN

	X ₀ OBJECT _{ACC} NICHT AUSSTEHEN KÖNNEN	X ₀ OBJECT _{ACC} NICHT AUSHALTEN KÖNNEN	X ₀ OBJECT _{ACC} NICHT LEIDEN KÖNNEN	X ₀ OBJECT _{ACC} NICHT ERTRAGEN KÖNNEN
Types (V)	303	70	246	166
Tokens (N)	718	150	672	397
<i>hapax</i> <i>legomena</i> (h1)	252	63	209	137
Baayen's \mathcal{P} (h1/N)	0.35097	0.42	0.31101	0.3451

Here, the resulting values for \mathcal{P} seem less clearly aligned with the expectations stemming from syntactic restrictions and the measures of colligational strength reported in Table 9 in the main text, which might have predicted that \mathcal{P} would be lowest for *ausstehen*. Instead, *ausstehen* is exceeded only by *aushalten* in terms of the number of distinct elements filling the object slot of the construction.

This discrepancy between the English and German results might partly be explained through further particular constructional preferences involving *aushalten*, *leiden* and *ertragen*. In the case of *aushalten*, although it had the proportionally greatest number of hapaxes, the occurrence of the pronouns *das* “that” and *es* “it” constituted much greater proportions of the tokens in comparison to *ausstehen*; with *leiden*, the same applies for *das*; with *ertragen*, the same applies for *es*, where more than one-third of all usages of *nicht ertragen können* showed *es* in the OBJECT_{ACC} slot, usually as a proleptic object used to introduce a following subordinate clause headed by *wenn* “if” or *dass* “that” (e.g. *die Geschichte von der Königin, die es nicht ertragen kann, dass ihre Stieftochter Schneewittchen schöner ist als sie* “the story of the queen who can’t stand it, that her stepdaughter Snow White is more beautiful than her; *Rhein-Zeitung* 20.06.2012, page 20). With *leiden*, the phrase *was ich/man nicht leiden kann* “what I/one can’t stand,” is exceptionally common, constituting more than 11% of all usages of *nicht leiden können* (*was* occurs as the object with *nicht ausstehen können* in only about 3% of tokens). *Leiden*, too, tends occur with *es* proleptically, like *ertragen*, whereas *es* with *ausstehen* more often occurs as a “genuine” pronominal object.

The fact, however, that *aushalten* exhibits the lowest degree of colligational strength with the X₀ OBJECT_{ACC} NICHT VERB_{INF} KÖNNEN construction but the highest degree of productivity as

measured by \mathcal{P} results in a generally negative correlation between colligational strength and productivity: Pearson's R between \mathcal{P} in Table A3 and German values in Table 9 of the main text for MINIMUM SENSITIVITY is -0.7088, and between the χ^2 likelihood ratio is -0.2777243. The German data thus confirm the same general tendency indicated in the English data, namely, that stronger collostructional strength between a lexeme and a construction correlates negatively with the productivity of the construction containing that collexeme.

The conservative conclusion to draw from this investigation of the productivity of lexically specific versions of the CAN NOT VERB_{INF} X₀ OBJECT_{ACC} construction is weak support for the hypothesis of a relationship between lower productivity and greater relative linguistic age. Somewhat surprisingly – from the point of view of this hypothesis – the differences in productivity between the German collexemes were smaller than in the English case, even though the German collexemes exhibited greater differences in degree of colligational strength and, in the case of *ausstehen*, stronger syntactic restrictions on the use of this verb (see Section 2 of the main text) are found than is the case for English *stand*. Overall, the reliability and utility of corpus-based measures of syntactic and morphological productivity as synchronic indicators of historical developments remains to be further demonstrated and qualified.

References

- Beam, C. Richard (Ed.). (2004–2011). *The Comprehensive Pennsylvania German Dictionary*. Volumes 1-12. With assistance of Joshua R. Brown, Jennifer L. Trout, and Dorothy Pozniko Beam. Center for Pennsylvania German Studies, Millersville University.
- Burger, H. (2007a). *Phraseologie. Eine Einführung am Beispiel des Deutschen* [Phraseology. An Introduction, Exemplified by German] (3rd ed.). Erich Schmidt.
- Burger, H. (Ed.). (2007b). *Phraseologie–ein internationales Handbuch der Sprach- und Kommunikationswissenschaft* [Phraseology – An International Handbook of Linguistics and Communications Science]. De Gruyter.
- Cotticelli-Kurras, P. (2007). Indogermanische Kasusyntax. Verwendung und Wandel des Akkusativs im Hethitischen [Indo-European case syntax. Usage and change of the accusative in Hittite]. *Studi Micenei ed Egeo-Anatolici* 49, 123–145.
- Covini, A. (2016–7). *Formazioni causative nelle lingue indoeuropee di più antica attestazione* [Causative Formations in the Oldest-attested Indo-European Languages] [Unpublished doctoral dissertation]. Universität zu Köln and Università di Siena.
- Donalies, E. (2009). *Basiswissen Deutsche Phraseologie* [Basics of German Phraseology]. A. Francke.
- Dunkel, G. (2014). *Lexikon der indogermanischen Partikeln und Pronominalstämme*. Band 1: *Einleitung, Terminologie, Lautgesetze, Adverbialendungen, Nominalsuffixe, Anhänge und Indices*. Band 2: *Lexikon*. [Lexicon of Indo-European Particles and Pronominal Stems. Volume 1: Introduction, Terminology, Sound Laws, Adverbial Endings, Nominal Suffixes, Appendices and Indices. Volume 2: Lexicon]. Carl Winter.
- Glare, P. G. W. (Ed.). (2012). *The Oxford Latin Dictionary* (2nd ed.). Oxford University Press.
- Grassmann, H. (1872 [1976]). *Wörterbuch zum Rigveda* [Dictionary to the *Rigveda*]. 5th edition. Harrassowitz.
- Firth, R. (1964). *Papers in Linguistics 1934-1951*. Oxford University Press.
- Hackstein, O. (2002). *Die Sprachform der homerischen Epen. Faktoren morphologischer Variabilität in literarischen Frühformen: Tradition, Sprachwandel, Sprachliche Anachronismen* [The Linguistic Form of the Homeric Epics. Factors of Morphological Variability in Early Forms of Literature: Tradition, Language Change, Linguistic Anachronisms]. Reichert.
- Harm, V. (2003). Zur semantischen Vorgeschichte von dt. *verstehen*, englisch *understand* und altgriechisch *ἐπίσταμαι* [On the semantic prehistory of German *verstehen*, English *understand*, and Ancient Greek *ἐπίσταμαι*]. *Historische Sprachforschung* 116, 108–127.
- Hornby, A. S., & Cowie, A. P. (1989). *Oxford Advanced Learner's Dictionary of Current English*. Oxford University Press.

- Jespersen, O. (1927). *A Modern English Grammar on Historical Principles*. Part III. Syntax. Second Volume. Carl Winter.
- Kroonen, G. J. (2010). On Gothic *iup* and the Germanic Directionals. *NOWELE : North-Western European Language Evolution*, 58–59, 367–379.
- Kulikov, L. (2012). Vedic preverbs as markers of valency-changing derivations. Transitivity and objecthood in Indo-European (Evidence from Old Indo-Aryan). *Studies in Language* 36(4), 721–746.
- Rix, H., Kümmel, M. J., Zehnder, T., Lipp, R. & Schirmer, B. (Eds.). (2001). *Lexikon der indogermanischen Verben: Die Wurzeln und ihre Primärstambildungen* [Lexicon of Indo-European Verbs: Roots and their Primary Stem-Formations] (2nd ed). Reichert.
- Kurz, J. (1958–). *Lexicon linguae palaeoslovenicae: Slovník jazyka staroslověnského*. Československá akademie Věd.
- Liddell, G. H., Scott, R., & Jones, H. S. (1925–1940). *A Greek English Lexicon*. Oxford University Press.
- Lühr, R. (2000). *Die Gedichte des Skalden Egill* [The Poems of the Skald Egill]. J. H. Röhl.
- Manning, C. D., & Schütze, H. (1999). *Foundations of Statistical Natural Language Processing*. MIT Press.
- Meiser, G. (1993). Zur Funktion des Nasalpräsens im Urindogermanischen [On the function of the nasal present in Proto-Indo-European]. In G. Meiser (Ed.), *Indogermanica et Italica. Festschrift für Helmut Rix zum 65 Geburtstag* (pp. 280–313). Innsbrucker Beiträge zur Sprachwissenschaft.
- Mottausch, K-H. (2015). Review of Scheungraber, Corinna: *Die Nasalpräsentien im Germanischen. Erbe und Innovation*. *Kratylos* 60, 95–102.
- Opfermann, A. (2016). *Univerbierung. Der passive Wortbildungsmechanismus* [Univerbation. The Passive Mechanism of Word-Formation]. Baar.
- Palm, C. (1995). *Phraseologie-eine Einführung* [Phraseology – An Introduction]. Gunter Narr Verlag.
- Prokosch, E. (1939). *A Comparative Germanic Grammar*. Linguistic Society of America.
- Ringe, D. (2017). *From Proto-Indo-European to Proto-Germanic*. (2nd ed.). Oxford University Press.
- Rix, H. (1995). Griechisch ἐπίσταμαι. Morphologie und Etymologie [Greek ἐπίσταμαι. Morphology and etymology]. In H. Hettrich., W. Hock, & P.-A. Mumm (Eds.), *Verba et Structurae. Festschrift für Klaus Strunk zum 65. Geburtstag* (pp. 237–247). Innsbrucker Beiträge zur Sprachwissenschaft.
- Sandell, R. (2015). *Productivity in Historical Linguistics: Computational Perspectives on Word-Formation in Ancient Greek and Sanskrit* [Doctoral dissertation, University of California, Los Angeles]. eScholarship University of California. <https://escholarship.org/uc/item/2z1476f8>
- Scheungraber, C. (2014). *Die Nasalpräsentien im Germanischen: Erbe und Innovation* [Nasal Presents in Germanic: Inheritance and Innovation]. Innsbrucker Beiträge zur Sprachwissenschaft.

- Schneider, C. (2011). Syntax und Wortarten der Lokalpartikeln des Ṛgveda X: *úd* [The syntax and part of Speech of the local particles of the Ṛgveda X: *úd*]. *Münchener Studien zur Sprachwissenschaft*, 65, 185–241.
- Schützeichel, R. (2012). *Althochdeutsches Wörterbuch* (7th ed.). de Gruyter.
- Villanueva Svensson, M. (2011). Anticausative-inchoative verbs in the northern Indo-European languages. *Historische Sprachforschung*, 124, 33–58.
- Zasada, A. (2020). *Die Nasalpräsentien im Vedischen* [Nasal Presents in Vedic] [Unpublished doctoral dissertation]. Ludwig-Maximilians-Universität, München.