

Between Words and Worlds

Jeong, Sunwoo**

How Linguistic Theories Inform

Whang, James***

Fictional Language Making

— A Case Study on the Film *HOPE**

ABSTRACT Speculative fiction has long provided a fruitful backdrop against which “conlang”-ing efforts have flourished, as evidenced by Tolkien’s Elvish, *Star Trek*’s Klingon, and many others. Although aspects of linguistic theory continue to guide conlangers, it remains open how exactly these theories, which are grounded in descriptive data on natural languages, hinder or help the conlanging endeavor, which often aims to push the imagination beyond the attested data. This study examines the creative process of making an alien language for Na, Hong-Jin’s 2026 film, *HOPE*. Offering the project as a case study, we propose that linguistic theories, especially those on linguistic typology and implicational hierarchies, can productively guide the creative process of conlanging, clarifying what types of phonemes, morphemes, morphosyntax, and sound-meaning associations are marked and why. To break the rules more creatively, one must first be fully acquainted with the rules of the game.

Keywords Linguistic Universals, Constructed Language (Conlang), Linguistic Typology, Implicational Hierarchy, Language in Fiction, Language in Film

* We thank director Na, Hong-Jin for reaching out and entrusting us with the conlang project for *HOPE*, and the film team at Forged Films, especially Ji-In Cheon, for facilitating collaborative discussions. We also thank the three anonymous reviewers and the editorial board of *Journal of Humanities* for helpful comments. Any remaining errors are our own.

** Associate Professor, Department of Linguistics, Seoul National University (First Author)

*** Associate Professor, Department of Linguistics, Seoul National University (Corresponding Author)

1. Introduction

One of the most compelling ways to create a fictional world is to endow it with its own language. Writers of speculative fiction such as Tolkien and Le Guin grasped this, whether by building fully elaborated languages (e.g., Elvish; Tolkien 1954) or by sketching radically alien ones as thought experiments (e.g., Nna Mmoy; Le Guin 2003). Such fully realized invented languages remain the exception rather than the norm, however. Building a conlang is an involved task, one that many creators do not have the time to tend to, in part because much of our linguistic knowledge is subconscious: while every one of us is a native speaker of at least one language, it takes training to consciously understand the systems and features that make up a language, fictional or otherwise. Perhaps this is why many recent conlangs, especially those figuring in large-scale projects such as *Arrival* (Villeneuve 2016) and *The Game of Thrones* (Benioff et al. 2011–2019), have been created by enlisting professional linguists and/or “conlangers” (see e.g., Peterson 2015; Coon 2020).

Although many agree that linguistic expertise is useful in creating a conlang, it remains open how exactly linguistic theories become relevant to different aspects of the conlanging process. After all, isn't part of the aim of a conlang to create a language that transcends the bounds of human languages? If anything goes in fiction, wouldn't linguistic expertise only impose rules and limitations, thereby hindering rather than helping our imagination?

The central thesis of this paper is that certain linguistic theories can play a pivotal role in guiding the imaginative processes behind conlanging. Specifically, we claim that theories of language universals, cross-linguistic

typology, and markedness (e.g., Jakobson 1968; Greenberg 1966; Haspelmath et al. 2005) can invigorate the conlanging endeavors by clarifying the parameters of linguistic variation, as well as the diverse sources and nature of linguistic universals and the markedness they imply. Put more simply, in order to find more creative ways of breaking the rules of a game, one must first know what those rules are, how they differ from each other, and what motivated them in the first place.

To build our case, we offer a case study. Having worked as professional linguists/conlangers on a recently released film project, *HOPE* (Na 2026), we show how considerations of diverse types of linguistic universals and markedness guided our process and helped us imbue our conlang, which we named /neʧneʃ/ (*lit.* ‘speaking-hearing’), with a compelling sense of “otherness.” We claim that a deeper understanding of linguistic markedness, together with a critical engagement with the relevant literature, enabled us to design the language so that it sounds just unfamiliar enough, while remaining enactable and approachable.

In the process, we aim to make two contributions. First, cutting across the usual methodological divide between generative and functional-typological approaches, we propose a taxonomy of linguistic universals that distinguishes absolute, implicational, and statistical universals, and also includes hybrid cases of statistical-implicational universals and perspective-dependent universals. Each type, we propose, is the source of a corresponding kind of “markedness.” Second, and more importantly, we demonstrate how this emerging taxonomy can function as a useful design resource, by working through each module of the grammar of /neʧneʃ/ and mixing and matching different types of markedness in creative directions. This dual aim sets our study apart from existing work at the intersection of linguistics

and conlanging. For instance, more practically oriented treatments such as Peterson (2015) provide helpful guidance on applying linguistic expertise to conlangs, but we focus here on offering a critical review of the literature on universals and distilling their insights into a useful analytical framework.

The paper is structured as follows. We begin by considering what the language for *HOPE* should sound and feel like, and how a handful of worldbuilding details can get us started. We then review the literature on linguistic universals and develop the taxonomy described above. With that scaffolding in place, we turn to /neɟneɟf/ itself, demonstrating in what ways it is marked, and thus rendered “other,” in its phonetics, phonology, morphosyntax, and sound-meaning associations.

2. Inhabiting the World

An invented language can enrich a fictional world only if it strikes a particular balance. It must feel genuinely other, yet remain accessible: as Mandala (2010, p. 30) observes, the estrangement that makes a conworld compelling fails if the language describing it becomes so different as to be inaccessible (see also Noletto 2024). Much like Goldilocks, then, we needed to find just the right sweet spot between “otherness” and familiarity.

How different are the aliens in *HOPE* from humans, and how should that difference register in their language? From the script provided by the director and a series of creative meetings, we formed the following impressions:

- (1) a. They are advanced sentient beings with broadly human-like vocal

tract anatomy; their technology for space travel and their understanding of the Universe exceed our own.

- b. They experience empathy, distress, and other familiar human-like emotions.
- c. Their society values loyalty and family affection. They have a royal family and a stratified social structure.

These traits point toward subtle rather than dramatic “othering.” (cf. See instances of more dramatically “othered” languages in section 3.1.) Because the aliens are broadly human in anatomy (1a) and in their cognitive and emotional lives (1b), their language must remain enactable by human actors and relatable to the audience; a radically alien system would misrepresent minds that are, in the relevant respects, close to our own. The stratified, royal society in (1c), by contrast, calls for a more salient linguistic signature, and it leaves one of the more discernable traits on /neɟneɟ/: an elaborate honorification system threading through both the nominal and the verbal morphology, which we develop in sec. 5.2.

3. Universals and Markedness:

“Otherness” from the Vantage Point of Linguistics

What makes human languages “human”? How do they distinguish humans from other communicative species, real or hypothetical? As we began our conlanging effort, one of our guiding questions was how to encode the “otherness” of the alien race in their language. A useful way to operationalize otherness is to first understand the “self,” or the origo:

the center that the “other” is other to, which here amounts to humans and human language. Having a firmer grasp of the core characteristics shared by all or most human languages should therefore help us craft a more compelling sense of otherness.

Studies of language typology and linguistic universals have long offered competing answers to the question of what is human in human language. The seminal conference at Dobbs Ferry, culminating in the volume edited by Greenberg (1966), helped funnel these into two main strains of modern inquiry. One is more top-down and deductive, grounded in the hypothesis of Universal Grammar (henceforth UG) and an innate linguistic faculty (Chomsky 1957; 1995); widely known as the generative approach, it is concerned with theorizing and confirming putative universals in the mental grammar of speakers. The other is bottom-up and inductive, grounded in documentations and typological surveys of existing, especially understudied, languages (Haspelmath et al. 2005; Croft 2002); widely known as the functional-typological approach, it is concerned with both the diversity and the emerging common cores of the world’s languages.

For our purposes, we begin establishing a slightly different classification of linguistic universals, one based less on methodological or theoretical commitments, and more on how a given universal may be formulated (as stand-alone generalizations or as conditional/relational ones), and whether and to what extent they admit exceptions. Let us now go over the emerging taxonomy, examining each type of universal in turn.

3.1. Absolute (“true”) universals

Are there features of language held by absolutely all human languages

without exception? The answer risks being trivial if those features are also shared by non-human sentient animals such as bees and bears. Communicative systems in general seem to use some modality (sounds, scents, physical movements) to convey a message; while this holds for human languages too, it is not a very interesting generalization.

In this respect, the “design features” proposed by Hockett (1963) offer a promising starting place. Hockett aimed to identify “true” universals, i.e., features shared by all human languages, by contrasting them with known animal communication. (2) provides a non-exhaustive list.

- (2) a. Vocal-auditory channel: the channel for all linguistic communication is vocal-auditory.
- b. Rapid-fading: all linguistic signals are evanescent.
- c. Arbitrariness: the relation between a meaningful element and its denotation is independent of any physical or geometrical resemblance between the two.
- d. Discreteness: the possible messages in any language constitute a discrete repertory rather than a continuous one.
- e. Semanticity: linguistic signals matter because there are associative ties between signal elements and features in the world; some linguistic forms have denotations.
- f. Openness (productivity): new linguistic messages are coined freely and easily.

For concreteness, let us consider a few. For one, all human languages are based on sequences of transient acoustic signals and are therefore rapid-fading (2b); other systems need not be, as those based on spours (e.g.,

mycelia) or trails (snails, mammals) sustain the same signal for a considerable time. Likewise, while the form-meaning mapping in human languages is generally arbitrary (2c; Saussure 1916), systems like bee-dancing may lack this feature, since it indexically or iconically represents e.g., the source of the nectar. The same holds for discreteness (2d): human languages consist of discrete units (phonemes, morphemes), whereas e.g., bee-dancing consists of continuous signals.

These features are in some sense domain-general: they do not target specific modules of grammar such as morphology or syntax, and many are grounded in fundamental human faculties (i.e., the nature of our cognitive, productive, and perceptual apparatus). Because even non-human characters, to be recognizable and relatable, tend to share these faculties, many conlangs observe them; Tolkien's Elvish languages, for instance, appear to observe all of them.

By contrast, whenever these features are violated, the repercussions are far-reaching and often seem to produce a strong sense of alienation. We believe this can be, and has been exploited strategically (though not always with an explicit awareness of the relevant features and the theories that support it): A conlang that violates design features can function as a worldbuilding or plot device to highlight the unknowability of another species, particularly in "first encounter" science fiction. For instance, Heptapod B in Ted Chiang's *The Story of Your Life* (Chiang 1998) appears to violate (2a), (2b), and likely (2c) and (2d), heightening the sense of mystique and impenetrability. Likewise, the main premise of Le Guin's *The Nna Mmoy Language* (Le Guin 2003) is to posit a conlang that violates (2e), semanticity. Specifically, the story envisages a radically context-dependent language, as described in the excerpt below.

“Each syllable is a word, but a word with no fixed, specific meaning, only a range of possible significances determined by the syllables that come before, after, or near it. A word in Nna Mmoy language has no denotation, but is a nucleus of potential connotations which may be activated, or created, by its context. Thus it would be possible to make a dictionary of Nna Mmoy only if the number of possible sentences were finite.” (Le Guin 2003, p. 145)

The net effect is that of a tantalizing mystery and lingering opacity the narrator feels for the Nna Mmoy people.¹

Beyond the design features, are there any other linguistic universals that can be called “absolute”? From the generative tradition, the property most often singled out is that of recursion (Chomsky 1957; Hauser et al. 2002): the capacity to generate an unlimited number of expressions from finite means, a notion that provides a theoretical anchor to ‘openness’ discussed in (2f).

Universals like Hockett’s design features and recursion are alike in being fundamental and domain-general: they characterize human language as a whole rather than any particular component of grammar. One may then ask whether there are absolute universals of a more domain-specific kind, exceptionless generalizations targeting a single module such as

1 It is worth mentioning that conlangs incorporating these more radical departures from human-like languages tend to appear only in partial forms or as thought experiments. Developing them into fully realized systems, especially for visual mediums like films, tends to be extremely challenging, if not impossible. Coon (2020) notes, for instance, that Heptapod B in the film *Arrival* (which she notes hasn’t been created by her, as she functioned as a consultant, not a conlanger) is not really a conlang, but rather, symbols inspired by paintings of an artist, Martine Bertrand.

phonology, morphology, or syntax. Here the terrain is murkier. The emerging consensus in linguistic typology seems to be that true, exceptionless universals are difficult to identify at the level of individual modules, given the diversity of the world's languages. Even patterns once thought dominant tend, on closer inspection, to resolve into strong tendencies rather than absolute laws.

With that caveat in mind, (3) presents some of the best, or least bad, candidates we could cull for individual modules of synchronic grammar. They are composited from different research traditions and do not form a natural class; we offer them less as established universals than as a working inventory that may be of interest to a conlanger.

- (3) a. Phonology: all languages have vowels and consonants. (Greenberg 1962, 1966)
- b. Morphology: all languages have at least some derivational morphology. (Bybee 1985)
- c. Syntax: all languages have nouns and verbs. (e.g., Evans and Osada 2005)
- d. Semantics (composition): all languages obey compositionality, i.e., the meaning of a whole is a function of the meanings of its parts. (Partee 2004)
- e. Semantics (conceptual categories; an example): every language has constituents whose semantic function is to express generalized quantifiers over the domain of discourse. (Barwise and Cooper 1981)

For some of these generalizations, full empirical confirmations may still be pending. For instance, the syntactic universal in (3c) was once taken for granted, but has since come under scrutiny (Croft 2002), with exceptions

offered such as Nootka (Hockett 1963, p. 4). Its ultimate status may depend on how grammatical categories are defined in the first place [see Cristofaro (2009) for a critical review].

Beyond the synchronic system, a further set of “absolute” universals may be proposed in the domain of language use (in context) and language change. We offer some candidates in (4).

- (4) a. Pragmatics: all languages are subject to (some version of) the Cooperative Principle and the Gricean maxims.
- b. Diachrony: all languages change.
- c. Sociolinguistics: all languages are simultaneously conventional and socially stratified.

Here again, some disclaimers may apply. For instance, Keenan (1976) noted that the Cooperative Principle (henceforth CP) and the Gricean maxims (4a) may not be operative in the same way in certain cultures/languages. As for (4b) and (4c), much more can be said regarding the systematicities that underlie language variation and change (e.g., cyclical or chain shifts as common patterns of change; stratification as common patterns of stylistic variation, etc.), which we refrain from elaborating further in the interest of space.

The candidates for absolute universals we’ve offered in (3) and (4) are by no means exhaustive, though they have been vetted enough to survive some level of empirical scrutiny. More than anything, they show how difficult it is to identify domain-specific “true” universals that are stand-alone (i.e., defined non-relationally) in formulation. This difficulty motivates the turn in sec. 3.2 to universals of a fundamentally relational kind.

What happens when these module-specific universals are violated?

We think some violations may pass unnoticed upon cursory inspection [e.g., (3e)]; while others can be as estranging as violating a design feature [e.g., (3a)]. We could find few existing conlangs, however, that targeted these universals specifically, which makes them an untapped resource for worldbuilding, opening doors to new possibilities.² For instance, could there be a language with only vowels (where much of its expressive load is carried by pitch and stress)? Or a language without generalized quantifiers? How about a language that *ain't gon' give rise to no* (case in point) variation, whether across groups (dialect) or individuals (idiolect)? More importantly, what would each of these languages tell us about their speakers and the ways in which they conceptualize the world? In each case, the linguistic anomaly and the imagined world inform each other.

3.2. Implicational universals

The universals in sec. 3.1 can all be stated in isolation: “All languages are, have, or do X.” But an influential strain of work formulates universals not in isolation but relationally, by making reference to a broader contrastive

2 One potential recent example that violates (3a), i.e., a “conlang” that lacks vowels and consonants, is the Eridian language in *Project Hail Mary* (Lord and Miller 2026; see also Weir 2021), which consists of sounds akin to musical notes, reflecting Eridians’ distinctive vocal apparatus. However, there’s a debate on whether the sounds included in the film (created by sound designers) can be construed as a conlang proper, as it lacks a fleshed-out grammar that goes beyond consistent mappings between certain musical sounds and words. As noted in fn. 1, this further highlights the practical difficulties involved in translating pre-conlang thought experiments (often entertained in written work) into fully realized conlangs. Looking ahead, perhaps collaborations between sound designers and linguists/conlangers can result in conlangs that are both sonically and systematically compelling, successfully instantiating violations of phonological “true” universals such as (3a), while also maintaining a developed grammar.

system. Intuitively, these take the form: “If a language has X, then it also has Y.” Such generalizations are called implicational universals.

A pioneering work in this vein has been offered by Jakobson (1968). He proposes that the emergence of linguistically meaningful sound contrasts (i.e., phonemes and distinctive features) follows a universally fixed order. The contrast between fricatives and stops, for instance, emerges only after the contrast between oral and nasal stops, i.e., the presence of a stop-fricative contrast presupposes the presence of an oral-nasal one. Strikingly, Jakobson shows that these implicational universals (the “laws of irreversible solidarity” (*solidarité irréversible*), to use his own terminology) manifest in three parallel domains: children’s acquisition, cross-linguistic typology, and aphasia. For instance, the stop-fricative case manifests in the following ways:

- (5) a. A language that has fricatives in its inventory also has stops.
- b. A child who has acquired fricatives as meaningful speech sounds has already acquired stops.
- c. An aphasic who has lost stops has already lost fricatives.

To provide a more specific example, (5a) can be instantiated as follows: a language which has [s] in its phoneme inventory will also have [t], though the reverse is not necessarily true.

A notion closely related to implicational universals is that of markedness. In the case above, we may say that fricatives are marked, or more precisely, more marked than stops. Jakobson (1968) notes several further hierarchies of this kind: back consonants are more marked than front ones (e.g., a language with [k] also has [p] and [t]); front rounded vowels are more

marked than front unrounded ones (e.g., a language with [ø] also has [e] and [o]), and so on.

If Jakobson established implicational universals in the domain of phonology, Greenberg (1963) established analogous generalizations in morphosyntax. Below are a few examples:

- (6) Universal 3. Languages with dominant VSO order are always prepositional.
- (7) Universal 13. If the nominal object always precedes the verb, then verb forms subordinate to the main verb also precede it.

Here and throughout, V, S, and O stand for verb, subject, and object, respectively. As with Jakobson's hierarchies, what is claimed to be universal in (6)–(7) is not a particular word order (cf. sec. 3.3), but rather the interdependencies among certain morphosyntactic settings within a language. Not all of Greenberg's generalizations are meant to be exceptionless like (6)–(7), however. Some hold only as strong statistical tendencies, as we will see in sec. 3.3.

At this point, a brief remark on the term “markedness” is in order, which we will begin using in more earnest. In its narrow, Jakobsonian sense, markedness is a fundamentally relational notion grounded in structural considerations: features that maximize contrast within a system are less marked and emerge first. Front rounded vowels, for instance, emerge only after front unrounded and back rounded vowels, not because they are harder to pronounce, but because unrounded front vowels maximize contrast with rounded back ones. Rounding the lips lengthens the overall vocal tract slightly, effectively pushing all vowels deeper to the

back and further away from unrounded front vowels.

Moving forward, however, we adopt a broader understanding of linguistic “markedness,” and use it as a cover term for whatever makes a form or pattern dispreferred or less expected, regardless of whether the markedness emerges relationally or not, and regardless of whether the markedness is due to production, perception, or processing difficulties.

Have implicational universals been sufficiently exploited to create “otherness” or “familiarity” in conlangs? Our survey is not exhaustive, but our initial impression suggests that they have not. More often, creators evaluate the markedness of a sound or a pattern in isolation, rather than markedness arising from the system as a whole. A possible counter-example is Dothraki (Peterson 2015; see also Benioff et al. 2011–2019 and Martin 1996), which lacks bilabial stops but has coronal and velar ones (thereby violating the Jakobsonian implicational hierarchy). While Peterson does not directly invoke the notion of implicational hierarchy, he does remark that the choice was made to “give the language a unique sound (addition by subtraction, as it were),” in addition to noting that he dislikes the phonaesthetic quality of bilabials (Peterson 2015: p. 88).

Implicational universals thus strike us as another fertile, untapped ground which conlangers can mine. For instance, could we imagine a language with the nasal stops [m], [n] but not their oral counterparts [p], [t]? With front rounded vowels [y], [ø] but not the unrounded [i], [e]? More ambitiously, could a language cluster all of its consonants and vowels at the back, with no front articulations at all? (This would be a very inefficient way to establish a contrastive system!) We expect all of these to sound “other” to varying degrees, even though the individual sounds (e.g., [m], [n], back consonants) are not in and of themselves marked. As linguists, we may also

want to motivate these putative violations: Perhaps the species' hearing capabilities are hyper-sensitive so that the contrastive pressures become effectively irrelevant. Or perhaps the species simply lacks lips. Here again, linguistic "markedness" and worldbuilding mutually inform each other. As we will see, /neɟneɟ/ draws directly on this strategy, in its back-weighted, fricative-heavy consonant inventory and its restrictions on labial stops (secs. 4.1, 5.1).

3.3. Statistical universals

We now turn to universals that are not, strictly speaking, absolute. In every domain, certain features, phonemes, or morphosyntactic processes are marked or uncommon. Whether this is grounded in human difficulties of production, perception, or processing, or is merely accidental (e.g., due to the vagaries of cultural history, language contact, etc.), the upshot, well established by typological studies, is that not all logical possibilities occur with equal frequency: Some are vanishingly rare, though not impossible, while others are extremely common. These tendencies can be exploited to create a sense of "otherness," and indeed have often been used to build practically usable conlangs. They can be stated as standalones, in the form of: "Most languages have/do (not) X." (8) provides an example in the domain of phonetics/phonology.

(8) Most languages do not use clicks or pulmonic ingressesives.

Click sounds are among the typologically rarest, occurring in only 10 of 567 languages (about 1.8%) in Maddieson's WALS sample (Maddieson 2005).

Pulmonic ingressives used phonemically, rather than as paralinguistic backchannels, are likewise vanishingly rare (Ladefoged and Maddieson 1996). Though these are extreme cases, we may construe frequency-based markedness more generally as a continuum. This enables us to characterize other rare sounds such as pharyngeal or lateral fricatives as being statistically/typologically marked, though not to the same degree as clicks.

Analogous statistical generalizations can be found in the domain of word order, as exemplified in (9).

- (9) Most languages follow SOV or SVO order.

SOV and SVO are the two most common orders, occupying 47.1% and 41.2% of the 1,052 languages investigated in Haspelmath et al. (2005). VSO and VOS are rarer, at 8% and 2.5%, and OVS and OSV rarest of all, at 0.9% and 0.4% (Dryer 2005).

Statistical universals can also be combined with implicational formulations (sec. 3.2), resulting in implicational “tendencies.” Greenberg’s universal in (10) provides a canonical case, concerning the same domain of word-order.

- (10) Universal 4. With overwhelmingly greater than chance frequency, languages with normal SOV order are postpositional. (Greenberg 1963)

Such statistical-implicational universals sit between the exceptionless relational universals of sec. 3.2 and the purely frequency-based ones of this section. We therefore list them as a distinct type in Table 1.

In conlanging, knowledge of these typological markedness patterns

has often been put to fruitful use, yielding languages that feel non-human and “other,” while remaining familiar and comprehensible enough.

Before closing this section, we will complicate our taxonomy just a step further. Though statistical universals are useful, what generates most effect in the practice of conlang might have less to do with a feature’s raw, statistical markedness, and more with relative, perspective-dependent markedness, for lack of a better term. Specifically, even a typologically unusual feature can feel unremarkable if it belongs to the audience’s native language or a prominent second language. Pharyngeal sounds, for example, are cross-linguistically rare, and may strike many Korean or anglophone listeners as “foreign.” Yet, to speakers of certain Semitic languages, they will sound entirely ordinary. We therefore treat perspective-dependent markedness as a further, derived case in Table 1. It will play a central role in our design choices to follow (sec. 4.1).

3.4. Summary of universals and markedness

Let us summarize. Human languages share certain key features, which we have called universals. We distinguished three core types, together with two derived cases, according to how each is formulated (stand-alone vs. relational) and whether and to what extent it admits exceptions. Table 1 presents the emerging taxonomy, pairing each universal with corresponding notions of markedness.

As captured by the final column, each type of universal can be paired with a corresponding kind of markedness, which is defined as a violation of the said universal.

Familiarity with these universals, and with the markedness they

[Table 1] A taxonomy of Universals and Markedness.

Type of Universal	Formulations	Corresponding markedness
Absolute	All languages have/do X.	Absolute
Implicational	If a language has/does Y, then it has/does X.	Implicational
Statistical	Most languages have/do X.	Statistical
Statistical-implicational	Most languages have/do X, if they have/do Y.	Statistical-implicational
Perspective-dependent	One’s native language and/or one’s salient L2/L3/⋯ have/do X.	Perspective-dependent

generate, helps clarify what makes human languages human, and in turn illuminates how a conlang can be made to feel both “other” and familiar. We also remarked on an asymmetry in how conlangers have drawn on them: Hockett’s design features and the statistical universals have been exploited more often (though the violation of the former is more often entertained as thought experiments rather than being integrated into a fully realized conlang), whereas domain-specific absolute universals and implicational universals remain comparatively untapped, offering a promising avenue for future creative work.

With this taxonomy in hand, we may now ask the following question regarding the design strategy of /neɕneɕ/: what kinds of markedness will do the brunt of the “othering” work, how, and where (if domain/module-specific)?³

3 Here, we’re making an implicit leap from “otherness” construed as a kind of typological markedness (i.e., violations of the universals outlined in Table 1) to “otherness” as perceived alienness, with the guiding assumption that the former will, to varying degrees, map onto the latter (intro to section 3). As reviewer 2 points out, however,

The worldbuilding that emerged from sec. 2, as well as practical considerations, provided a useful first-pass orientation. To begin with, we decided not to violate any “absolute universals” in /neɟneɟf/, with the possible exception of arbitrariness (we postpone its discussion until sec. 6).⁴ This led us to direct our attention to implicational, statistical, and perspective-dependent markedness. In sections that follow, we show how /neɟneɟf/ incorporates diverse combinations of these “non-absolute” markedness in its phonetics and phonology (secs. 4.1 and 5.1) and its morphosyntax (secs. 5.2 and 5.3).

actual audience perception data would be needed to more definitively establish this link. It also remains open what specific instantiations of markedness types summarized in Table. 1 will result in greater vs. subtler intuitions of perceived alienness. Though we have proposed, in broad strokes, that absolute markedness will be perceived as most alien (section 3.1), and in a similar vein, may conjecture that implicational markedness will rank higher than statistical markedness in their effect size, these claims are not granular enough to predict how much perceived alienness will be achieved by a specific design choice (and combinations thereof) for a given audience.

- 4 Regarding this decision, reviewer 3 wonders why more radically creative linguistic choices have not been taken, along the line of Heptapod B (Chiang 1998) or other conceivable violations of absolute universals. In addition to the world-building (section 2) and practical constraints that were imposed (the language needed to be performable by human actors), we highlight again the difference between conceptualizing a radical language in a written work (a sort of ‘pre-conlang’) and implementing it as a fully fleshed-out conlang in visual/collaborative media such as films. (See also fn. 1 and fn. 2.) We hope more experimentations in the former (some promising avenues were proposed in section 3), and more cross-disciplinary collaborations in the latter, will help ease the technical/budgetary restrictions and bridge the gap between imagination and implementation. We also hope that the subsequent sections demonstrate that even without incorporating absolute markedness, interesting creative choices can be made to build sufficient “otherness” in language.

4. The building blocks: Otherness in the sounds and morphemes of /neɕneɕʃ/

As we saw in sec. 3.1, all human languages can produce an unlimited number of sentences (and the associated concepts) from a finite set of ingredients, by virtue of recursion. /neɕneɕʃ/ is no exception in this way. Within this broad scaffolding, however, both the finite ingredients of /neɕneɕʃ/ (in our case, phonemes and morphemes) and the combinatorial rules that enable recursion (i.e., its morphosyntax) carry marked features that make the language feel “other.” This section examines how such estrangement is achieved at the level of ingredients.

4.1. The sounds

In principle, the sound inventory of a conlang is unbounded: one could draw on anything from the whistle of a teakettle to the phonemes of Alsatian French. Practical and worldbuilding considerations rein this in, however. How can the sound system of /neɕneɕʃ/ be made just strange and “other” enough, while remaining vocalizable by human actors and relatable enough for the audience? Our answer applies, in turn, four types of markedness summarized in sec. 3.4: perspective-dependent markedness to shape the overall palette, then statistical, implicational, and statistical-implicational markedness to fine-tune the inventory.

We began with perspective-dependent markedness, suggesting to the director a handful of natural languages as references: Icelandic, Saami, Inuktitut, and Mongolian. These were chosen because their sound systems would strike most Korean and anglophone audiences as relatively unfamiliar

(that is, they are perspective-dependently marked) and because each displays at least a few typologically unusual phonemes or sound patterns. They also came close, phonaesthetically, to the ambience the director wanted: cold, ancient, and mystical. Positing a loose sound-symbolic association helped narrow the search early: we predicted, through informal probing of our own phonaesthetic intuitions, that fricative-heavy languages such as Icelandic would more reliably evoke coldness,⁵ and this was borne out in our meetings with the director.

Having established that the director preferred the sound qualities of a particular dialect of Mongolian, we turned to fine-tuning the inventory, with Mongolian serving only as a point of departure. Here we were guided by statistical markedness, including its implicational variety. Tables 2 and 3 give the finalized phoneme inventory of /neɟneɟ/.

We created a sense of “otherness” in the inventory in three ways. The first was to include phonemes that are statistically and typologically marked (i.e., violate statistical markedness). The most salient are the lateral fricative /ɭ/ and the bilabial fricative /β/. In PHOIBLE 2.0, a cross-linguistic database of phonological inventories (Moran and McCloy 2019), /ɭ/ occurs in only 48 of 3,020 inventories (about 1.6%); it does occur in Mongolian, and we think it is part of what gives the language its distinctive character. Likewise, Maddieson (2005) notes that bilabial fricatives are less common than bilabial stops and nasals. Affricates, while not as statistically infrequent, are also

5 We make no claim here that there is an inherent link between fricatives and coldness. We leave open the possibility of a historical accident, where languages spoken by prehistoric communities that settled in these parts simply happened to have/develop a fricative-heavy inventory, which we over time have come to associate with the cold climate in which these are spoken. The key takeaway, however, is that whatever the reason, this intuition seems largely shared.

[Table 2] Consonant inventory.

	Labial		Coronal		Velar	
	plain	palatalized	plain	palatalized	plain	palatalized
Nasal	m	mʲ	n	nʲ	ŋ	
Stop	p	pʲ	t	tʲ	g	gʲ
Affricate			ts	tʃ		
Fricative	(f β)		s	ʃ	x	ç
			ʁ			
Liquid			r			
Approximant	w	wʲ	j			

[Table 3] Vowel inventory.

	Front	Central	Back
High	i, iː	u, uː	u, uː
Mid	e, eː	ʌ, ʌː	o, oː
Low		a, aː	

more marked than stops, and the preponderance of fricatives, especially in back articulatory regions such as the velum, is also marked. Among the vowels, high central unrounded vowels like /u/, though not vanishingly rare, are also relatively uncommon compared to high front and back vowels.

The second strategy was to incorporate a statistical-implicational markedness in the inventory. Specifically, we included voiced velar stops while excluding their voiceless counterparts. Voiced stops are cross-linguistically more marked than voiceless ones, such that the presence of a voiced stop typically presupposes that of its voiceless counterpart. Arabic offers a well-known counterexample, with /b/ but no /p/. Aerody-

namically, however, voicing becomes progressively harder to sustain as the place of articulation moves backward in the oral cavity: the farther back the closure, the smaller the cavity behind it, and the sooner air pressure builds up to inhibit the vocal fold vibration necessary for voicing. Voiced velar stops are accordingly more marked than their coronal and labial counterparts, and their inclusion was a deliberate strategy for introducing a further sense of otherness.

The third strategy, related to the second, was to incorporate a weak, frequency-based version of full implicational markedness. Recall (sec. 3.2) that the presence of fricatives (more marked) presupposes that of their matching oral and nasal stop counterparts (less marked), and that the presence of back stops (more marked) presupposes that of their front counterparts. As a way of violating these implicational universals, which yields a deliberately wonky contrastive system, we initially considered doing away with labial oral stops altogether, while keeping coronal and velar stops, along with labial fricatives.

A strong version of this vision could not be maintained, as the proper names the director had given us, the names of alien characters, already contained labial stops. We therefore adopted a weaker implementation. The inventory retains labial oral and nasal stops, but we restricted their distribution so that they surface only in certain positions (sec. 5.1), whereas fricatives are not similarly constrained. We also modulated the relative frequency of fricatives (and affricates) compared to stops, so that the language sounds more heavily weighted toward the former: this was achieved by planting more fricatives than stops into the key inflectional morphemes (sec. 5.2.2) and by keeping the preference in mind during other morpheme selection (sec. 4.2).

4.2. The lexicon

Contrastive phonemes are not themselves meaning-bearing units. For that, we need morphemes. Having fixed the sound inventory (sec. 4.1) and the basic phonotactics (sec. 5.1), we generated the licit morphemes of /neçneŋ/ through a semi-randomized procedure. At its center is a “root generator,” a Python script that takes the language’s phoneme inventory, together with the phonotactic constraints that govern licit sound sequences, and returns well-formed roots. We restricted roots to a single syllable, of the shape CV(V)(C); the fuller syllable template CVVC(C) is reached only through affixation, and the second coda consonant never belongs to a root (sec. 5.1). The generator is included in the documentation linked in the Appendix.

We randomized this process not only for efficiency but, more importantly, to keep our own native intuitions out of the lexicon. As native speakers of Korean and English, if we were to hand-pick every morpheme, we would inevitably smuggle in native phonotactic preferences and frequency patterns, thereby eroding the unfamiliarity we were trying to build. Generating roots at random, within explicitly non-native phonotactic constraints and from a non-native inventory, keeps those biases out, so that the intended otherness survives not only in which sounds occur but in how often they do. [See also Pearson (2020) for a similar approach.]

This largely automatic derivation was supplemented by more hands-on decisions. Among the candidate roots the script returned, we chose ones that maximized sound-symbolic associations (sec. 6). To satisfy phonaesthetic considerations, we also filtered the output to honor the ‘fricative over stop’ weighting described in sec. 4.1. Certain high-frequency

inflectional and derivational morphemes, along with keywords that serve as emotional anchors in the script, were chosen even more deliberately, through multiple consultations with the film team.

Our decision to restrict roots to a single syllable went hand in hand with a commitment to a highly systematic, robust derivational morphology. We aimed to maximize the economy and recyclability of morphemes, so that a single root could generate a whole family of related concepts through derivation. This degree of economy, and the dense interconnection of words and concepts it produces, is unlikely to be found in natural languages to the same extent, since natural languages are continually reshaped by impersonal pressures like language contact and sound change. The near-categorical economy of our root system is thus a further, non-phonological source of markedness, distinct from the inventory-level markedness of sec. 4.1. It also encodes part of the implicit mythology we assumed for the aliens and the world they inhabit. We return to this systematicity in the derivational morphology of sec. 5.2.

5. The glue: Otherness in the phonology and morphosyntax of /neçnetʃ/

Otherness and markedness can emerge not only in the ingredients a language uses, but in how those ingredients are combined. This section turns to the phonology and morphosyntax of /neçnetʃ/, and shows how the same subtle “otherness” is built into its combinatorial rules in ways we hope reflect the minds and cultures of the aliens.

5.1. Putting the sounds together: Otherness in phonology

A sequence of sounds can feel foreign not only when an individual sound is marked or non-native, but also when familiar sounds appear in marked or non-native positions. English, for instance, has /ŋ/ in its inventory (as in “-ing”), but disallows it word-initially. A word beginning with /ŋ/ therefore sounds odd to English speakers, even though the sound itself is familiar (word-initial /ŋ/ is also typologically marked). Phonotactic constraints of this kind, governing which sound sequences are licit, provide a fruitful backdrop for channeling markedness into a sense of “otherness.”

In establishing the phonotactics of /neɟneɟ/, we drew primarily on perspective-dependent, statistical-typological, and implicational markedness. The licit syllable structure for /neɟneɟ/ is CVVC(C), where the second coda consonant can arise only through affixation. The language permits just two complex codas, both word-final: /xt/ and /st/. This is marked in two respects.

First, the presence of codas is typologically more marked than their absence, and complex codas are more marked than simple ones. Implicationaly, a language that allows CVC also allows CV but not conversely, and one that allows CVCC also allows CVC but not conversely. For a Korean audience, complex codas already sound “other,” since Korean phonology disallows them. And although English permits complex codas, the particular profile of /neɟneɟ/, admitting /xt/ while disallowing more expected clusters such as /lt/ or /nt/, will register as “other” to Anglophone listeners too, since English does the reverse.

Second, the available complex codas /xt/ and /st/ violate an implicational universal tied to the Sonority Sequencing Principle (Clemens 1990;

henceforth SSP). Coda clusters that fall in sonority, such as a liquid-stop sequence like /lt/, are less marked than clusters that do not; implicationaly, a language that allows /xt/ should also allow /lt/, but not conversely. /neɟneɟ/ inverts this, admitting only the less optimal clusters and forbidding the more optimal ones. While the coda /st/ violates SSP, it is not cross-linguistically infrequent (i.e., is not statistically marked). Its markedness in /neɟneɟ/ thus emerges less from its presence per se, and more from its presence despite the absence of e.g., coda /nt/ (implicational markedness). Coda /xt/ by contrast, is both statistically and implicationaly marked.

The phonotactics of /neɟneɟ/ is wonky in a further respect: it disallows word-initial labial stops, even though labial stops are implicationaly less marked than their back counterparts, and word-initial position is among the least marked places for them to occur (being maximally distinctive there).

In both cases, otherness is achieved not only through the *presence* of an unusual sequence but through the *absence* of an expected one: e.g., the missing /lt/ (given /xt/) and the missing word-initial /pa-/.

This only scratches the surface of /neɟneɟ/ phonology. Much more could be said about its suprasegmentals (stress, intonation) and its vowel phonology. (For instance, /neɟneɟ/ has backness harmony.) In the interest of space, however, we have concentrated on aspects that are marked, and thus do the brunt of the “othering” work. We refer the interested reader to the full documentation linked in the Appendix.

5.2. Putting morphemes together: Otherness in morphology

How does /neɟneɟ/ build words from morphemes? Its word formation is unusual at both the inflectional and derivational level. In each, the

otherness is largely statistical and typological in kind: The processes themselves are attested in human languages, but their particular shape, productivity, and pervasiveness in /neçnef/ are marked.

5.2.1. Derivational morphology

As adumbrated in sec. 4.2, /neçnef/ has a rich and markedly systematic derivational morphology. We realized this in two ways: through robust compounding, and through a templatic morphology that systematically links the parts of speech (verb, noun, adjective) which share a conceptual field.

First, we relied heavily on compounding, letting certain compounds reduce to a manageable length through reduction rules. Thus, “spacecraft” is /teem-so/ (literally, time/space-vehicle), “warship” /e-xar-so/ (war-vehicle), “wormhole” /çiira-num/ (space-knot), and “moment” /teem-ğats/ (time-piece, shortened from /teem-ağat + nats/). Compounding is not confined to nouns: “to believe” is /çor-sax/ (literally, to hold the Universe), and “to hope” is /çor-gats/ (to see the Universe). Compounding is not itself a marked process, but its sheer pervasiveness in /neçnef/ is unusual relative to Korean or English, and, like the root system of sec. 4.2, it directly reflects the mythology we assumed for a species that prizes economy and interconnectedness.

Second, we established a productive templatic morphology, relating concepts that differ mainly in part of speech. The schema is given in (11).

(11) A templatic morphology of /neçnef/ (a sample template: /tVf/)

- a. Verbs (standardly roots): back vowels, e.g., “to chill” /tuɸ/
- b. Adjectives: central vowels, e.g., “cold” /tuɸ/
- c. Nouns: front vowels, e.g., “chill, coldness” /tiɸ/

Treating /neçneʃ/ as a verb-centric language (sec. 6), we derive deverbal adjectives from the verb root by shifting its vowels from back to central, and deverbal or deadjectival nouns by shifting back/central vowels to front. Semantically, a deverbal adjective standardly denotes the resultative state of applying the verb, and a derived noun, the associated abstract property: so “to chill” (make cold) is /tuʃ/, “cold” is /tuʃ/, and “chill, coldness” is /tiʃ/. Verbs are the default roots, but a denominal verb can also be formed by the same rule, meaning “to be (a) N.” For instance, /çuuro/ “to be fated/destined” (literally, to be the Universe) derives from /çiira/ “Universe.”

This kind of templatic, non-concatenative morphology is typologically rare. Cross-linguistically, concatenative morphology (prefixation and suffixation) is far more common than nonlinear morphology (Greenberg 1960; Bickel and Nichols 2005), though Semitic languages like Arabic and Hebrew provide famous exceptions. And even where templatic morphology occurs, it is usually confined to a particular domain. The productivity and pervasiveness we assume for it in /neçneʃ/ are therefore statistically and typologically marked, and we expect them to function as a further source of “otherness.”

Finally, /neçneʃ/ rounds out its derivational morphology with a range of prefixes, affixes, and even infixes. The deverbal noun for the “act of V-ing” is formed with the prefix /e-/, as in /e-çire/ “fate/destiny” (the act of being fated). Emphasis is marked by the infix /-pa-/, placed after the first (often only) syllable of the root: /ɾʌt-ast/ “firstly” becomes /ɾʌt-p(a)-ast/ “truly.”

5.2.2. Inflectional morphology

Not all languages have robust inflectional morphology, even though all have at least some derivational ones (Bybee 1985). Grammatical relations may be marked by morphemes, by word order, or by some mix of the two. /neçnef/ leans toward the morphological end, and one feature of its verbal inflection is especially marked.

In the nominal domain, the system is fairly standard. Case marking is agglutinative and does little “othering” work (the full paradigm is given in the documentation linked in the Appendix). Its one notable wrinkle is honorification, which runs as a fully separate case paradigm extending even to more marked cases such as the genitive and locative.

The verbal domain is where the bulk of the othering work is realized. Cross-linguistically, inflectional affixes on the verb tend to occur in a fixed order, relative to the stem: Aspect is the closest, then tense, then mood, with person-number affixes being farthest out (Bybee 2009). Exceptions to this are rare, but exist, making it a strong statistical universal.

/neçnef/ violates this strong statistical universal rather conspicuously: the mood marker is a prefix, while voice, modality, aspect, and tense follow the stem as suffixes, as shown in (12).

(12) Order of inflectional affixes on the verb in /neçnef/

MOOD – VERB. RT – (VOICE) – (MODALITY) – (ASPECT) – TENSE

Table 4 provides the mood-marking prefixes, which, like the nominal case system, are stratified by honorification into plain, formal, and royal registers.

The prefixal mood marker is exemplified in (13), a question formed with

[Table 4] Mood marking by speech style.

Mood	Plain	Formal	Royal
Indicative (ind)	-	nə(ʃ)-	ʒə(ʃ)-
Interrogative (int)	ən-	ʃən-	ʃə(ʒ)-
Imperative (imp)	ə(t)-	nə(t)-	ʒə(t)-
Hortative (hort)	əm-	nəm-	ʒəm-
Subjunctive (sbjv)	ə-	nən-	ʒən-

the interrogative prefix /ən-/. The honorific stratification surfaces in (14), where a servant addresses a royal and the indicative carries its honorific prefix /nəʃ-/.

- (13)

niç ən-ɔʃpa-çurt-a?

you INT-run-can-PRES

‘Are you able to run?’
- (14)

xas nəʃ-ost-asa ses-jaʃ eçire-rim gertu-ja rina-mi.

I IND.HON- he-GEN destiny-ACC Gh’ertu-GEN name-DAT

change-FUT

‘I shall change his destiny (in the name of Gh’ertu).’

In sum, /neçneʃ/ shows statistical markedness in the order of its verbal inflections. It also shows pervasive honorific stratification across both the nominal (argument) and verbal (addressee) domains, functioning as linguistic signatures of the highly stratified, royal society described in sec. 2.

5.3. Putting words together: Otherness in syntax

Because /neçneʃ/ carries so much grammatical information in its

inflection, word order bears little of the burden in signaling grammatical relations. The language thus permits scrambling on information-structural grounds. We nonetheless posited a default order, SVO, which is not itself unusual, being the second most common order after SOV (Dryer 2005). The fuller canonical order is given in (15).

(15) Canonical word order of /neçneŋʃ/

S(NOM) V PP P(LOC) R(DAT) O(ACC)

The default order is not where the “otherness” lies. It comes from the syntax violating a statistical-implicational pattern. As hinted in sec. 3.3, the ordering of elements across constituents typically tends to align in one direction, motivating the broad split between head-initial and head-final languages. For instance, SVO languages tend to have prepositions and postnominal relative clauses, whereas SOV languages tend to have postpositions and prenominal clauses. /neçneŋʃ/, however, refuses to align in either direction. While its default word order is SVO, its phrasal modifications precede the noun, as in (16), where “that leads to water” modifies “road” from the left. (This is a constructed sentence that does not occur in the actual film.)

- (16) metxar aɔʃ-xʌn xoç ʃees aɟut-a.
 that go-PTCP to water be.road-PRES
 ‘That is a road that leads to water.’

The upshot of this mixed head-directionality is perspective-dependent markedness: The syntax of /neçneŋʃ/ looks “other” for both Korean

audiences, used to a clearly head-final language, and anglophone ones, used to a clearly head-initial language.

6. Otherness in sound-meaning mappings

The markedness we have drawn on so far have been statistical, implicational, or perspective-dependent. Does /neçnetʃ/ touch on the absolute universals at all? The one design feature it comes close to violating is arbitrariness. This is because we built in as many sound-symbolic form-meaning associations as we could.

Tense marking is one example. The past affixes (/–ana/ great past, /–ata/ past) use stops, whereas the future affixes (/–asa/ future, /–axa/ great future) use fricatives, mirroring the articulatory contrast between the two: what has passed has clear beginnings and endings, what is to come is open and continuous.

The templatic morphology of sec. 5.2.1 is sound-symbolic in a similar vein. Taking the aliens to understand the essence of things as change and movement, we made verbs the base form and derived nouns and adjectives from them. Importantly, we used front vowels for nouns, because “things” are construed as ephemeral and surface-level, and back vowels for verbs, because actions and states in flux are construed as feeling truer and deeper.

Sound symbolism is of course attested in natural languages as well (Ohala 1994; Kawahara et al. 2018), but the consensus is that it stays confined to local domains (e.g., diminutives), without overriding arbitrariness at large. [See also Kawahara (2020) for a recent overview, Sullivan and Kang (2025), Winter and Perlman (2021), and Jeong (2017) regarding sound symbolisms in the domain of proper

names, size adjectives, and sociolinguistic variables, respectively.] /neɕneɕf/ likewise remains arbitrary on the whole. But it is still set apart from natural languages in the sheer pervasiveness and systematicity of its sound-symbolic associations. In short, /neɕneɕf/ challenges arbitrariness without completely abandoning it.

7. Discussion and conclusion

In the first half of this paper, we reviewed linguistic theories on universals and markedness. We assessed how they have implicitly or explicitly guided existing conlangs, and distilled the discussion into a new taxonomy. In the process, we also highlighted certain types of markedness which are under-explored in conlangs (domain-specific absolute markedness and implicational markedness), and suggested promising avenues for future experimentation. In the second half, we offered a case study of how the emerging taxonomy informed our own work, the alien language /neɕneɕf/, for the film *HOPE* (Na 2026).

The case study offers two contributions. First, it showcases that knowledge of implicational and statistical markedness lets one build a sound system that feels unfamiliar and “other” without having to resort to outlandish phonemes. Likewise, it demonstrates that knowledge of implicational and perspective-dependent markedness lets one design strange yet systematic morphosyntactic patterns, which can serve as rewarding Easter eggs for attentive audiences.

Through our investigation, we sought to show that while part of the pleasure of conlanging is pushing past the rules that govern human

languages, the theories that describe those rules are exactly what let us transgress them deliberately. To break the rules more creatively, and to know just when, how, and why, one must first understand the rules (or in some cases the tendencies), themselves. Having made a case that linguistic theories have much to offer to conlangers and creators of speculative fiction, we “hope” (pun intended) that our creative output, made in collaboration with the film team, proves compelling to audiences and future conlangers alike.

8. Appendix

A fuller documentation of the conlang /neɟneɟ/ can be found in:
<https://github.com/sunwooj/HOPE-conlang>

References

- Barwise, Jon, and Robin Cooper. 1981. “Generalized Quantifiers and Natural Language.” *Linguistics and Philosophy* 4 (2): 159–219.
- Benioff, David, D. B. Weiss, George R. R. Martin, Carolyn Strauss, Frank Doelger, Bernadette Caulfield, Bryan Cogman, Miguel Sapochnik, and David Nutter (executive producers). 2011–2019. *Game of Thrones* [TV series]. Home Box Office (HBO).
- Bickel, Balthasar, and Johanna Nichols. 2005. “Fusion of Selected Inflectional Formatives.” In *The World Atlas of Language Structures*, edited by Martin Haspelmath, Matthew S. Dryer, David Gil, and Bernard Comrie, 86–89. Oxford: Oxford University Press.
- Bybee, Joan L. 1985. *Morphology: A Study of the Relation between Meaning and Form*. Amsterdam and Philadelphia: John Benjamins.
- Bybee, Joan. 2009. “Language Universals and Usage-Based Theory.” In *Language*

- Universals*, edited by Morten H. Christiansen, Chris Collins, and Shimon Edelman, 17–39. Oxford: Oxford University Press.
- Chiang, Ted. 1998. "Story of Your Life." In *Starlight 2*, edited by Patrick Nielsen Hayden. New York: Tor Books.
- Chomsky, Noam. 1957. *Syntactic Structures*. The Hague: Mouton & Co.
- Chomsky, Noam. 1995. *The Minimalist Program*. Cambridge, MA: MIT Press.
- Clements, George N. 1990. "The Role of the Sonority Cycle in Core Syllabification." *Papers in Laboratory Phonology 1*: 441–79.
- Coon, Jessica. 2020. "The Linguistics of Arrival: Heptapods, Field Linguistics, and Universal Grammar." In *Language Invention in Linguistics Pedagogy*, edited by Jeffrey Punske, Nathan Sanders, and Amy V. Fountain, 32–48. Oxford: Oxford University Press.
- Cristofaro, Sonia. 2009. "Grammatical Categories and Relations: Universality vs. Language-Specificity and Construction-Specificity." *Language and Linguistics Compass* 3 (1): 441–79.
- Croft, William. 2002. *Typology and Universals*, 2nd ed. Cambridge: Cambridge University Press.
- Dryer, Matthew S. 2005. "Order of Subject, Object and Verb." In *The World Atlas of Language Structures*, edited by Martin Haspelmath, Matthew S. Dryer, David Gil, and Bernard Comrie, 330–33. Oxford: Oxford University Press.
- Evans, Nicholas, and Toshiki Osada. 2005. "Mundari: The Myth of a Language without Word Classes." *Linguistic Typology* 9 (3): 351–90.
- Greenberg, Joseph H. 1960. "A Quantitative Approach to the Morphological Typology of Language." *International Journal of American Linguistics* 26 (3): 178–94.
- Greenberg, Joseph H. 1962. "Is the Vowel-Consonant Dichotomy Universal?" *Word* 18 (1–3): 73–81.
- Greenberg, Joseph H. 1963. "Some Universals of Grammar with Particular Reference to the Order of Meaningful Elements." In *Universals of Language*, edited by Joseph H. Greenberg, 73–113. Cambridge, MA: MIT Press.
- Greenberg, Joseph H., 2nd ed. 1966. *Universals of Language*. Cambridge, MA: MIT Press.
- Haspelmath, Martin, Matthew S. Dryer, David Gil, and Bernard Comrie, eds. 2005. *The World Atlas of Language Structures*. Oxford: Oxford University Press.
- Hauser, Marc D., Noam Chomsky, and W. Tecumseh Fitch. 2002. "The Faculty of Language: What Is It, Who Has It, and How Did It Evolve?" *Science* 298 (5598): 1569–79.
- Hockett, Charles F. 1963. "The Problem of Universals in Language." In *Universals of Language*, edited by Joseph H. Greenberg, 1–29. Cambridge, MA: MIT Press.
- Jakobson, Roman. 1968. *Child Language, Aphasia and Phonological Universals*. The

- Hague: Mouton.
- Jeong, Sunwoo. 2017. "Iconization of Sociolinguistic Variables: The Case of Archetypal Female Characters in Classic Hollywood Cinema," In *Dimensions of Iconicity*, edited by Angelika Zirker, Matthias Bauer, Olga Fischer, and Christina Ljungberg, 263–286. Amsterdam: John Benjamins Publishing Company.
- Kawahara, Shigeto. 2020. "Sound Symbolism and Theoretical Phonology." *Language and Linguistics Compass* 14 (8): e12372.
- Kawahara, Shigeto, Atsushi Noto, and Gakuji Kumagai. 2018. "Sound Symbolic Patterns in Pokémon Names." *Phonetica* 75 (3): 219–244.
- Keenan, Elinor Ochs. 1976. "The Universality of Conversational Postulates." *Language in Society* 5 (1): 67–80.
- Ladefoged, Peter, and Ian Maddieson. 1996. *The Sounds of the World's Languages*. Oxford: Blackwell.
- Le Guin, Ursula K. 2003. "The Nna Mmoy Language." In *Changing Planes*. San Diego: Harcourt.
- Lord, Phil, and Christopher Miller, (directors). 2026. *Project Hail Mary*. Amazon MGM Studios, Premiered at the London World Premiere, March 9, 2026.
- Maddieson, Ian. 2005. "Presence of Uncommon Consonants." In *The World Atlas of Language Structures*, edited by Martin Haspelmath, Matthew S. Dryer, David Gil, and Bernard Comrie, 82–85. Oxford: Oxford University Press.
- Mandala, Susan. 2010. *Language in Science Fiction and Fantasy: The Question of Style*. London: Continuum.
- Martin, George R. R. 1996. *A Game of Thrones*. New York: Bantam Books.
- Moran, Steven, and Daniel McCloy. 2019. *PHOIBLE 2.0*. Jena: Max Planck Institute for the Science of Human History.
- Na, Hong-Jin, (director). 2026. *HOPE*. Plus M Entertainment and Forged Films. Premiered in Competition at the Cannes Film Festival, May 17, 2026.
- Noletto, Israel A. C. 2024. *Fictional Languages in Science Fiction Literature: Stylistic Explorations*. London and New York: Routledge.
- Ohala, John J. 1994. "The Biological Bases of Sound Symbolism." In *Sound Symbolism*, edited by Leanne Hinton, Johanna Nichols, and John J. Ohala, 325–47. Cambridge: Cambridge University Press.
- Partee, Barbara H. 2004. *Compositionality in Formal Semantics: Selected Papers*. Oxford: Blackwell Publishing.
- Pearson, Matt. 2020. "Using Language Invention to Teach Typology and Cross-Linguistic Universals." In *Language Invention in Linguistics Pedagogy*, edited by Jeffrey Punske, Nathan Sanders, and Amy V. Fountain, 86–106. Oxford: Oxford University Press.
- Peterson, David J. 2015. *The Art of Language Invention: From Horse-Lords to Dark Elves*,

- the Words Behind World-Building*. New York: Penguin Books.
- Saussure, Ferdinand de. 1916. *Cours de Linguistique Générale*. Edited by Charles Bally and Albert Sechehaye. Lausanne and Paris: Payot.
- Sullivan, Lisa, and Yoonjung Kang. 2025. "Sound Symbolism in Korean Names: The Interplay between Cross-linguistic and Language-specific Patterns." *Korean Linguistics* 21 (1): 61-100.
- Tolkien, J. R. R. 1954. *The Lord of the Rings*. 3 vols, London: George Allen & Unwin, 1954-55.
- Villeneuve, Denise, (director). 2016. *Arrival*. Paramount Pictures, FilmNation Entertainment, and Lava Bear Films. Premiered at the Venice Film Festival, September 1, 2016.
- Weir, Andy. 2021. *Project Hail Mary*. New York: Ballantine Books.
- Winter, Bodo, and Marcus Perlman, 2021. "Size Sound Symbolism in the English Lexicon." *Glossa: A Journal of General Linguistics* 6 (1): 1-13.

원고 접수일: 2026년 7월 11일, 심사완료일: 2026년 8월 8일, 게재 확정일: 2026년 8월 10일

초록

말과 세계 사이에서

정선우*

언어 이론은 어떻게 인공어 창작에
기여할 수 있는가?

제임스 황**

— 영화 〈호프〉 사례를 중심으로

〈반지의 제왕〉의 엘프어, 〈스타트렉〉의 클링온 등과 같은 사례가 보여주듯, 사변 소설(speculative fiction) 관련 창작물들은 오랫동안 인공 언어(conlang) 제작에 창의적인 토대를 제공해왔다. 언어학 이론들은 그동안 여러 방면에서 인공 언어 제작자(conlanger)들을 인도해 왔지만, 실존하는 자연 언어를 설명하고자 하는 이론들이, 존재하지 않는 언어에 대한 상상력을 요구하는 인공 언어 제작 작업에 정확히 어떠한 기여를 하는지에 대한 평가 및 분석은 아직 언어학적 관점에서 체계적으로 이루어지지 않았다. 본고는 영화 〈호프〉(나홍진 2026)에서 사용되는 인공 언어 창작을 이와 같은 관점에서 고찰한다. 〈호프〉 외계어의 구상 과정을 구체적인 사례로 제시하면서, 언어학 이론, 특히 언어유형론과 함축적 위계(implicational hierarchies) 관련 이론이 인공 언어 구상에 생산적인 기여를 할 수 있음을 밝힌다. 특히 어떤 유형의 음소, 형태소, 통사규칙 등이 유표적(markedness)인지, 그것들이 어떻게, 왜 유표적인지를 고찰하는 과정이 언어 창작과 언어 이론 사이에서 상호적인 통찰을 제공할 수 있음을 보여주면서, 규칙을 보다 창의적으로 깨뜨리기 위

* 서울대학교 언어학과 부교수, 제1저자

** 서울대학교 언어학과 부교수, 교신저자

해서는 먼저 게임의 규칙을 온전히 이해하고 있어야 함을 논증한다.

주제어 언어 보편성, 인공 언어, 외계어, 언어 유형론, 함축적 위계, 소설
속 인공 언어, 영화 속 인공 언어

