The linguistics of *Arrival*:
Heptapods, field linguistics, and Universal Grammar[1]
Jessica Coon
McGill University

1. The Story of Your Life

I first read the short science-fiction work *The Story of Your Life* by Ted Chiang—on which the film *Arrival* is based—in the summer of 1999, just after it came out. It was the year after I had graduated from high school and before I would start college. While I now can't be sure whether it was really the first time I had heard the word "linguist", I am confident it was the first time I encountered "field linguistics." The story came to me seemingly by accident, one chapter away from a friend's mother's story in a *Year's Best Sci-fi* volume that happened to be lying on a coffee table in their living room one summer afternoon in rural Oregon. From a Heptapod's point of view, though, I think it is safe to say that nothing should be considered truly an accident.

***

Two years later, just after finishing my sophomore year in college, I arrived in Mexico for my first summer of linguistic fieldwork. I flew to Mexico City and took a bus twenty hours south to the beautiful mountain town of San Cristóbal de las Casas. Once there I was to meet up with my undergraduate linguistics professor, a renowned fieldworker, anthropological linguist, and Mayan languages expert named John Haviland. Haviland gave me a tour of town, installed me in a guestroom in his house, and after a few days instructed me to repack my bags. The next morning, before dawn, we started out in his truck down winding mountain roads into the hot jungle lowlands, into Ch'ol country. With us that morning was a young Ch'ol-speaking woman who was working on a linguistics MA thesis there in San Cristóbal, along with her two small nieces, all returning home.

It is clear to me in retrospect that during that seven- or eight-hour trip I should have been asking more questions: What did I need to know about Ch'ol culture? Were there things I should definitely do? Or definitely not do? What should my goals be for that summer? How did one get started doing fieldwork? And importantly: how should I get back to San Cristóbal? Instead, I spent much of time staring out the window silently, in a combination of awe at the beautiful scenery—pine forests giving way to thick jungle, cornfields perched on cliff sides, people of all ages carrying impossibly large bundles of firewood and corn on their backs along the highway—and panic at the realization of what I had signed myself up for. My silence may also have been due in part to my recognition of just how poor my Spanish really was, as Haviland and his student chatted incomprehensibly, and perhaps a concerted focus on my part to not get carsick.

We arrived sometime in the midafternoon heat at a Ch'ol-speaking Mayan village. Children were sent to find adults, and eventually we were ushered into a small home with a thatched roof, a packed earth floor, and wood still smoldering next to the *comal* in the elevated cooking fire (I remember Haviland remarked that back up in the highlands, the cooking fires are always directly on the ground to help heat the house; here in the jungle the extra heat would be unwelcome). My surprised host family to-be included the MA student's brother, his wife, and four children: the girls from the truck ride, and their two brothers.

It is possible that they had not been warned about our arrival because it simply is not easy to get messages into this village, which at the time had a single satellite phone that rarely worked. Or perhaps they were not alerted on purpose, because it would be easier to say no to this strange request without me standing in their home, looking bewildered and (maybe, slightly) pitiful. Whatever the case, after a negotiation I mostly didn't understand, Haviland got ready to drive back to the city. Overwhelmed, with no Ch'ol to speak of, a handful of introductory linguistics courses on my transcript, and my courage quickly slipping away, I asked my professor to remind me again what exactly I was supposed to be doing there that summer. "Make some friends," he said casually, "learn some Chol." Then he got back in his truck and drove away.

2. Arrival
I have to imagine that Dr. Louise Banks, the fictional field linguist and protagonist of the 2016 motion picture *Arrival*, knows the feeling (for more on this feeling, and some tips on how to lessen it, see Monica Macaulay's excellent 2004 article: 'Training Linguistics Students for the Realities of Fieldwork'). A similar mixture of panic, excitement, and self-doubt must have begun to settle in as she was rushed by military helicopter from her comfortable university office to the site of an enormous alien spaceship. Once there, she is tasked with deciphering the language of the recently-arrived Heptapods—a task daunting enough to shake anyone's confidence.

Though the *Arrival* filmmakers consulted me on many aspects of the linguistics in the film, the comparison between Dr. Banks' situation and my own that summer in Mexico is, in truth, hardly fair. For one thing, my job in Chiapas was to learn about the grammar of Ch'ol, a language spoken by around 200,000 indigenous Maya people in southern Mexico. Today there are about thirty different languages belonging to the Mayan family in Mesoamerica (the exact number depends on what counts as a dialect, and what as a separate language). Taken all together, today speakers of Mayan languages total more than six million. This group of languages is called a "family" because the modern languages are descended from a common ancestral language, Proto-Mayan, partially reconstructed by historical linguists and proposed to have been spoken roughly four thousand years ago. Though there was relatively little work on Ch'ol itself when I started out, the Mayan language family as a whole is one of the best documented language families in the Americas (Aissen et al. 2017).

Dr. Banks' job, on the other hand, was to decipher Heptapod, a language spoken by at least two giant aliens from somewhere deep in outer space (or maybe at least twenty-four giant aliens if one counts the twelve different spaceships that landed around Earth, and assumes that each ship is staffed by two creatures). Whereas I had a body of research on related Mayan languages to read up on, a recent MA thesis by a native speaker of Ch'ol, and even an old Ch'ol–Spanish dictionary produced by missionaries, there is of course no "Learn Heptapod" book for Dr. Banks,

no work on related languages (that we know of), and even the best language-learning apps or translation software won't help with Heptapod.

Our work environments were very different too. While I still had a lot to learn about Ch'ol culture, I also had a very welcoming and patient host family to live and work with. They quickly integrated me into the family as something like a useless older sister. I was barely able to sleep in a hammock, open coconuts with a machete, or get my clothes clean in the river—skills which any competent eight-year-old should have mastered. My tortillas never turned out very well, but I discovered that grinding corn was a good arm workout, requiring endurance but not much skill, and I could help out by grabbing groceries on my way back from my weekly trips to Salto de Agua, the nearest town with payphones and an internet cafe. I learned that "going for a walk" was not something that women were expected to do alone, but I could volunteer to trek to the river to carry back water from the spring (though while my young host sisters could carry large buckets up and down over the hills without spilling a drop, I required containers that sealed shut in order to avoid an extra bath). Exactly what I was doing learning Ch'ol was not immediately clear to anyone, but everyone rose to the challenge of teaching me their language, patiently answering my requests for translations, judgments, and slow repetitions.

Dr. Banks, on the other hand, had a military tent, a cot, and two giant, seven-limbed aliens in a spaceship. There were other important differences as well. Whereas I had a knowledgeable undergraduate research supervisor in a not-so-distant city (which I eventually did learn how to travel back to), she had military generals yelling at her to hurry up. While my undergraduate honours thesis felt like a very daunting task at the time, Dr. Banks faces impending world war if she doesn't get things right. I also had one more significant advantage, though it was not obvious to me at the time: I had Universal Grammar on my side.

3. "Universal" Grammar
Here on Earth, language sets humans apart from all other species. Human babies—remarkably bad at basic tasks like feeding themselves, tying their shoes, and adding sums of numbers—effortlessly learn any language (or languages) to which they get sufficient exposure. While children make mistakes along the path of acquisition, even these mistakes follow certain patterns and developmental trends. By the age of five, nearly every child has mastered a complex system of grammar that organizes sounds into words and words into sentences. Beginning from a very early age, kids can produce and comprehend an infinite number of novel utterances—a feat that anyone who has tried to learn a new language as an adult can appreciate.

What linguists call *Universal Grammar* is the innate human capacity for language: core principles that all human languages share. Though languages show an apparently high degree of variation—the grammar of English is different from the grammar of Japanese, which is different from the grammar of Inuktitut—linguists have discovered that languages vary in limited and constrained ways. In fact, languages tend to follow certain *recipes* in their grammars (see Mark Baker's *The Atoms of Language* for an accessible introduction to some of these recipes). The syntax of Japanese, for instance, looks in many important ways like the syntax of Quechua, an unrelated language indigenous to the Andes mountains in South America. Both languages show a basic subject–object–verb (SOV) word order, and in both, direct objects are marked with

accusative case, as shown by the sentences in (1) and (2).[2]

(1)   Mariya papa-**ta**   ranti-chka-n.
               Maria   potato-ACC buy-PROG-3SG
               'Maria is buying potatoes.'   (Quechua)

(2)   Maria-wa zyagaimo-**o**   ka-ttei-ru.
               Maria-TOP potato-ACC   buy-PROG-NONPAST
               'Maria is buying potatoes.'   (Japanese)

Just as the order of verbs and objects are the reverse of what we find in English, "adpositions"—little words like *in*, *on*, and *at*—in both Japanese and Quechua also follow the nouns with which they combine. In English we call them prepositions, while in Japanese and Quechua they are *post*positions, as shown by the examples in (3) and (4).

(3)   wasi-kuna-**pi**
               house-PL-in
               'in the houses'   (Quechua)

(4)   ie-**ni**
               house-in
               'in a house   (Japanese)

Overwhelmingly, these properties tend to cluster together across human languages: if the verb precedes the object, the language will have prepositions; if the verb follows the object, the language will have postpositions. This is known as Greenberg's Universal #4—one of a list of language "universals" (or in many cases, tendencies), documented by the linguist Joseph Greenberg. These patterns lead linguists to hypothesize that variation among human languages is constrained to certain *parameters*. Children acquiring language have a head-start because their brains are hard-wired with at least some of the basic building blocks of language. Given sufficient exposure to a particular language, these parameters get fixed one way or another: once a kid learns that the verb precedes the object, that kid can then be reasonably confident that an adposition will precede its noun.

In another example, Niuean, a Polynesian language spoken on the island of Niue, shares a number of grammatical properties with Q'anjob'al, a Mayan language related to Ch'ol spoken in the highlands across the border in Guatemala. Both Niuean and Q'anjob'al have a relatively rare basic order of verb–subject–object (VSO), found in fewer than 10% of the world's languages (and following Greenberg's Universal #3, these and other VSO languages are languages with prepositions). Both languages also show what is known as an ergative-absolutive pattern of alignment, in which transitive subjects pattern differently from intransitive subjects; this pattern is found in roughly one quarter of the world's languages (Dixon 1979; Coon et al. 2017). In Niuean in (5), the transitive subject appears with *he* (not found on intransitive subjects), while in Q'anjob'al in (6) the transitive subject triggers a special prefix *s-* on the verb, also not found with intransitive subjects.

---

[2] Abbreviations in example sentences are as follows: ABS – absolutive; AP – antipassive; ACC – accusative; ERG – ergative; NOM – nominative; NONPAST – non-past tense; PL – plural; PFV – perfective aspect; PROG – progressive; PST – past tense; SG – singular; SUBJ – subject; TOP – topic.

(5)   Ne   kai [s **he**   pusi] [o ia   e   moa].
PST eat   ERG cat   that ABS bird
'The cat ate the bird.'   (Niuean)

(6)   Max **s**-tzok'   [s naq winaq] [o te'   si'   ].
PFV 3ERG-chop   the   man   the   wood
'The man chopped the wood.'   (Q'anjob'al)

Interestingly, both languages also sometimes permit objects to appear without articles (in indefinite non-specific contexts). When this happens, the ergative marking disappears, and the order changes to VOS, as shown in the examples in (7) and (8).

(7)   Ne   inu   [o kofo kono] [s e   Mele].
PST drink   bitter coffee   ABS Mary
'Mary drank bitter coffee.'   (Niuean)

(8)   Max tsok'-wi [o si'   ] [s naq winaq   ].
PFV chop-AP   wood   the   man
'The man chopped wood.'   (Q'anjob'al)

Do these three characteristics—VSO order, ergative marking, and VOS order with article-less objects—pattern together for a reason? Or is this an accident? If it's not an accident, *why* should these relatively unique properties go together? These are the kinds of research questions that linguists are interested in.

Linguists working on understudied *human* languages benefit in different ways from the same head-start that human babies have. A linguist who learns that a subject of a transitive sentence in Ch'ol triggers a special prefix on the verb is not surprised to also learn that possessors trigger an identical prefix on a possessed noun—because exactly this pattern (specifically, a syncretism between ergative and possessive morphology) is found in unrelated languages around the world. A linguist working on a VSO language like Q'anjob'al expects that question words like 'what' and 'who' must appear at the beginning of a sentence, like in English, while a linguist working on an SOV language like Japanese is unsurprised to learn that question words may remain in their base positions.

But when it comes to describing the grammar of the newly arrived Heptapods, even the most seemingly basic human language distinction, like the difference between "nouns" and "verbs", or between a statement and a question, is no longer a given. Linguists who coined the term "Universal Grammar" had only the universe of *human beings*—not Heptapods—in mind. They weren't thinking that far ahead.

3. Heptapod A and Heptapod B
In *Arrival*, following the plot of *Story of Your Life*, the seven-limbed Heptapods have two different languages: a spoken language (dubbed "Heptapod A" in the book) and a written language ("Heptapod B"). Heptapod A, we learn in more detail in the short story, does not exactly follow the patterns of human languages. But as Dr. Banks notes, it is also not wildly different:

> *It didn't follow the pattern of human languages, as expected, but it was comprehensible so far: free word order, even to the extent that there was no preferred order for the clauses in a conditional statement, in defiance of a human language "universal." It also appeared that the heptapods had no objection to many levels of center-embedding of clauses, something that quickly defeated humans. Peculiar, but not impenetrable* (Louise Banks, in Ted Chiang, *The Story of Your Life*).

Having flexible word order is not uncommon in languages of the world—but even given word order variation and flexibility, human languages nonetheless maintain certain consistencies. As the quote from Dr. Banks notes, one such property (Greenberg's Universal #14) is that the antecedent of a conditional (the *if*-clause) will always precede its consequent (the result). Despite the fact that English and Japanese have reverse orders when it comes to verbs and their objects and adpositions and their nouns, the order of clauses in a conditional is identical:

(9)　　Ame-ga　　fut-ta-ra,　　watasi-wa　　ie-ni　　　i-ru.
　　　　rain-NOM　fall-SUBJ-if　I-TOP　　　　house-in　stay-NONPAST
　　　　'If it rains, I'll stay home.'　　　　　　　　　　　　　　　　　(Japanese)

Heptapod A also apparently permits rampant center-embedding, another feature that *human* languages universally tend to avoid. Consider the sentences in (10). The sentence in (10b) adds a relative clause modifying the subject, 'the cat', to the basic sentence in (10a)—embedding a clause in the middle of another clause. However, if we try to add another instance of embedding, as in (10c), things become dramatically worse, and (10d) is basically word salad (indicated in linguistics by a *).

(10)　　a.　The cat fell.
　　　　b.　The cat [that the dog chased] fell.
　　　　c.　??The cat [that the dog [that the mouse scared] chased] fell.
　　　　d.　*The cat [that the dog [that the mouse [that the bird saw] scared] chased] fell.

The interesting thing about the degraded nature of English (10c) and (10d) is that this can't be easily pinned to a rule of the syntax specifically; as (10b) illustrates, embedding a clause inside another is not in and of itself a problem. Instead, the problem has been claimed to be one of *processing*. Our human short-term memory has a difficult time storing up the subjects and then later matching them to their disjoint predicates. In (10d), for example, one has to wade through three other clauses before *the cat* can be associated with its predicate, *fell*. Heptapods are apparently unfazed by this extra tax on memory load—perhaps unsurprisingly, given what we learn about their general cognitive capacities.

Indeed, the devout *Arrival* fan will also not find it surprising that the order of conditional clauses may be reversible in Heptapod. Chiang has clearly done his homework here. While Heptapod A violates human language norms, it does so in an expected way based on what we know of the Hetapods' special cognitive abilities. Chomsky's (1993) "Strong Minimalist Hypothesis", formulated for human language, is that language is an *optimal solution* to interface conditions—that is, to human conceptual and articulatory or sensory-motor needs. Though linguists are still working to understand exactly what these needs are, and what an optimal solution to them would

be, if it eventually turns out that all we need to do to predict properties of Heptapod is to modulate our theories appropriately for Hetapods' particular conceptual and articulatory abilities, then perhaps the term "Universal Grammar" may not be so far off the mark after all.

While Dr. Banks makes progress with the spoken Heptapod A, her real focus turns to Heptapod B. For one thing, unlike aliens in many sci-fi films, Heptapods do not have humanoid vocal tracts, and the sounds they make—created in the film by splicing together various animal calls with the help of my phonetician colleague Morgan Sonderegger, and described in the story as sounding "vaguely like that of a wet dog shaking the water out of its fur"—are not reproducible by humans. Indeed, as Dr. Banks notes, we can't even be sure our human ears would be able to pick out which sounds are meaningful. Heptapod B is also preferred over Heptapod A because Dr. Banks recognizes the importance of being able to interact directly with her language consultants. Colonel Weber initially approaches Banks to do the translation work by simply listening to audio recordings. In the story, she is having none of it:

> *But the only way to learn an unknown language is to interact with a native speaker, and by that I mean asking questions, holding a conversation, that sort of thing. Without that, it's simply not possible. So if you want to learn the aliens' language, someone with training in field linguistics—whether it's me or someone else—will have to talk with an alien. Recordings alone aren't sufficient* (Dr. Banks, in Ted Chiang, *The Story of Your Life*).

Once inside the shell, her attempts to reproduce the alien sounds with her human vocal tract predictably fail. While she is able to make recordings and play them back, she has much more luck working with the writing system.

The written Heptapod B is described by Chiang as being "semasiographic"—a symbolic way of communicating information without a direct tie to phonetic speech. While human writing systems differ from one another in important ways, they are all based in some way on the spoken language. Some human writing systems are alphabetic, with symbols representing individual sounds, while others are syllabic, in which symbols represent entire syllables. Still others are logographic, in which a symbol may represent an entire word or concept. Many languages of the world do not have writing systems at all; writing is a useful tool, but not a central part of the innate human cognitive capacity for language. What all human writing systems *do* have in common, is that the writing system is based on the spoken language (or perhaps, on an earlier version of it, as English spelling demonstrates).

We learn that Heptapod B, on the other hand, has no connection at all to Heptapod A. In the film, the Heptapod "semagrams" are beautiful circular swirly symbols; they look something like stains made from a coffee mug, but appear to move like smoke through the air, constantly changing. In the story, they are described by Chiang as resembling "fanciful praying mantids drawn in a cursive style, all clinging to each other to form an Escheresque lattice, each slightly different in stance."

Why have two separate languages in the first place? As the linguist-physicist duo speculate in both story and film, the Heptapods would similarly wonder why we humans aren't taking better advantage of these two distinct media. While spoken languages are constrained in time, written

languages need not be. Dr. Banks muses:

> *For them, speech was a bottleneck because it required that one word follow another sequentially. With writing, on the other hand, every mark on a page was visible simultaneously. Why constrain writing with a glottographic straitjacket, demanding that it be just as sequential as speech? It would never occur to them. Semasiographic writing naturally took advantage of the page's two-dimensionality; instead of doling out morphemes one at a time, it offered an entire page full of them all at once* (Dr. Banks, in Ted Chiang, *The Story of Your Life*).

While some people are disappointed to learn that I did not create the language for *Arrival*, often they are even unhappier to learn that it is not really a language at all. Instead, the symbols in the movie are based on the beautiful paintings of Montreal-based artist Martine Bertrand. While the filmmakers went to great lengths to achieve consistency across different scenes, and even created a small manual of roughly one hundred symbols used in different parts of the film, one cannot learn Heptapod B the way one can learn Klingon or Na'vi. (Nor can one sell merchandise promising to translate any English phrase into Heptapod, as one vendor hoped to solicit my expertise for.) And given the consequences of learning Heptapod B, as Dr. Banks does in the story, creating such a language would be no small feat.

Indeed, both the film and Chiang's original work are short on details of the grammar of Heptapod B. We do learn in the story that some elements of Heptapod B are "uniquely two-dimensional"—the curvature and thickness of a line, the manner of undulation, or the relative sizes, distances, and orientations of the meaningful elements all play important roles in the grammar. But neither Chiang nor the filmmakers can really be blamed for a lack of detail here. In fact, it isn't clear that even Dr. Banks could fully articulate how she has come to learn this language. Movie viewers watch a montage in the movie of her staying up late nights, furiously decoding semagrams, the wall plastered with symbols and scribbles (my scribbles, at the request of the set designers); concern is expressed for her mental and physical well-being as she begins to learn the language. In the short story, Louise notes that: "the semagrams seemed to be something more than language; they were almost like mandalas. I found myself in a meditative state, contemplating the way in which premises and conclusions were interchangeable." In both, the language-learning experience is clearly a bit surreal.

4. Linguistic Relativity

It is the uncertainty about how an alien language might differ from human language—and whether and how we humans might be able to learn such a language—that makes the premise of *The Story of Your Life~Arrival* so thought-provoking. The plot draws heavily on the Sapir-Whorf Hypothesis (left implicit in the short story, but discussed explicitly in the screen adaptation), also known as Linguistic Relativity. Most famously attributed to Benjamin Lee Whorf, Linguistic Relativity is the hypothesis that the language we speak directly affects how we view the world. According to Whorf, speakers of different languages have correspondingly different thoughts.

Whorf was a chemical engineer who studied linguistics with Edward Sapir at Yale University during the 1930s while still keeping his day job at the Hartford Fire Insurance Company. During this time, he also began his controversial work on the Uto-Aztecan language Hopi, which would

come to play a big role in his promotion of Linguistic Relativity. He wrote:

> *I find it gratuitous to assume that a Hopi who knows only the Hopi language*
> *and the cultural ideas of his own society has the same notions, often supposed*
> *to be intuitions, of time and space that we have, and that are generally*
> *assumed to be universal. In particular, he has no general notion or intuition of*
> *time as a smooth flowing continuum in which everything in the universe*
> *proceeds at an equal rate, out of a future, through a present, into a past...*
> (Benjamin Lee Whorf, 'An American Indian Model of the Universe').

According to Whorf, the Hopi people viewed time differently from English speakers, and this was a direct result of the nature of their language; Whorf claimed that the Hopi language had no way to directly refer to the present, past, or future—or to the passing of time at all. Whorf also famously spread the idea that Alaskan Yupik people's apparently expanded vocabulary for different types of snow meant that they must also think more precisely about snow than the average English speaker (Whorf 1940).

Whorf's linguistic claims about the Hopi language have since been discredited (e.g. Malotki 1983), and in his 1991 essay 'The great Eskimo vocabulary hoax,' Geoffrey Pullum, citing work by Laura Martin, notes that any English-speaking skier has as many words for snow as speakers of Yupik are claimed to have (think *snow*, *slush*, *sleet*, *powder, blizzard…*). Indeed, Pullum argues that the disputed size of the Yupik lexicon for snow is no more interesting than the fact that professional typesetters know more names of fonts than the lay typist; a fact that is barely noteworthy, and certainly not headline news. Among human languages—which appear to follow the same underlying principles, and differ in interesting but ultimately constrained ways—the strong version of the Sapir-Whorf Hypothesis, Linguistic Determinism, has been argued to be not only wrong, but also dangerous. As Whorf's quote above illustrates, it provides an easy rationale to exoticize people who speak different languages.

This is not to say that all people view the world in the same way, just that our view of the world is not strictly determined by the language we speak. In his book *The Language Hoax: Why the World Looks the Same in Any Language*, John H. McWhorter offers an accessible look into some of the debunked myths surrounding Linguistic Relativity, as well a survey of the more recent and interesting scientific work on subtle correlations between language and thought. In the end, however, McWhorter concludes that language is not the best place to look for differences among humans:

> *If you want to learn about how humans differ, study cultures. However, if you*
> *want insight as to what makes all humans worldwide the same, beyond*
> *genetics, there are few better places to start than how language works* (John
> H. McWhorter, *The Language Hoax: Why the World Looks the Same in Any*
> *Language*).

Indeed, based on the deep commonalities among human languages, Noam Chomsky has famously stated that a visiting Martian (or perhaps Heptapods) would view all human speech as essentially dialects of the same language.

While some linguists have complained about the spotlight given to the Sapir-Whorf hypothesis

in the film, in the end, we have to remember that it is science fiction—"fiction" being the operative word—and like all good science fiction, it does make one think: how could an alien's linguistic system differ from our own? What correlations might there be between their language and their cognitive system? And if their language were so different as to change our way of thinking… would we be able to learn it at all?

Personally, though I enjoyed my time working with the filmmakers, I think there are bigger things for Louise Banks and the rest of us linguists to complain about. Take Colonel Weber's attempt at flattery early in the film, when he tells Dr. Banks that she is at the top of everyone's list when it came to *translation* (groan), the physicist Dr. Donnely's attempt at flattery when he tells her that she really thinks like a scientist (duh), or Banks' opening-scene lecture about how Portuguese is different from other Romance languages (huh?). Though I marked these in red in multiple versions of the script, and offered to write Dr. Banks a more engaging introductory linguistics lecture for the spaceships to interrupt, the filmmakers gently explained to me that in the end, linguists are not Hollywood's main audience.

My interactions with the filmmakers did also force me to think: how *would* we get to work deciphering an alien language when the time comes? How hard would this be? While our knowledge of human language will only get us so far, the tools we have developed for linguistic fieldwork and analysis will be critical. At least I hope this is true, because if aliens have just showed up and someone at the FBI is googling "alien linguistics", my name comes up pretty high on the list.

5. Linguistic fieldwork and language diversity
In the movie *Arrival*, Dr. Banks recognizes that the constraints and patterns linguists know to be found in human languages may be of no help in her new fieldwork situation. She nonetheless approaches the daunting task of deciphering the Heptapod language as any good fieldworker would. Inside the Heptapod shell, she is the first to take off her protective spacesuit and approach the glass divide. While theoretical linguists are interested in the abstract properties of language— the formal system that allows us to put sounds together to make words, and words together to make sentences—access to that system is not direct, but must be accomplished by careful work with native speakers of the language in question. As Dr. Banks knows, establishing a positive working relationship is the first step in any successful data-gathering activity.

Louise Banks also knows that progress doesn't happen overnight. Despite the urgent orders of military generals to get to the point—*why are they here?*—Dr. Banks insists that she must start with the basics. Even seemingly benign concepts, like asking a question or grounding an event in space or time, may have no direct correlate in Heptapod. But if there is hope of deciphering Heptapod at all, it should at least be *compositional*: that more complex concepts are formed in systematic ways from smaller units of meaning, as they are in human languages. Jumping straight to an exciting complex sentence before understanding the smaller parts from which it is constructed is a first-order error learned in any linguistic field methods class (and then often learned again the hard way when one is actually in the field).

Finally, Banks recognizes that misunderstandings are virtually guaranteed, and that one can't be too careful drawing conclusions from freshly collected data from an unfamiliar language. My

first summer in Chiapas, I was especially interested in Ch'ol's VOS word order—a basic order found in 3% or fewer of the world's languages. Interestingly, Ch'ol is like Q'anjob'al insofar as VOS order is found when the object has no article, and VSO occurs when it has one. But unlike Q'anjob'al, it is not uncommon to find articleless nouns in Ch'ol; articles in Ch'ol are not strictly required for definite interpretations, the way they are in English or Q'anjob'al. Also unlike Q'anjob'al, the ergative marking does not disappear in a Ch'ol VOS sentence, as shown in (11). The Q'anjob'al VOS sentence from above is repeated in (12) for comparison.

(11)   Tyi   **i**-kuchu   [$_O$ si'   ] [$_S$ jiñi   x'ixik   ].
              PFV   ERG3-carry   wood   the   woman
              'The woman carried wood.'   (Ch'ol)
(12)   Max   tsok'-wi [$_O$ si'   ] [$_S$ naq   winaq   ].
              PFV   chop-AP   wood   the   man
              'The man chopped wood.'   (Q'anjob'al)

A number of other factors have been claimed to interact with word order alternations in Mayan: animacy, definiteness, and specificity, as well as the relative ranking of the subject and object with respect to some combination of these properties. In my first summer of linguistic fieldwork, I naively thought I would jump in by trying to systematically isolate these factors, asking my host family for judgments on the sentences I created.

—Can I say *tyi ikuchu jiñi si' x'ixik*? What about *tyi ikuchu si' x'ixik*? And could I say *Tyi ikuchu jiñi si' jiñi x'ixik*?

The answer to every question, over and over again, to my great puzzlement, was "yes". At some point, after a few sessions of this, it occurred to me to rephrase my question.

—Wait… can <u>you</u> say *tyi ikuchu jiñi si' x'ixik*?
—No! I would definitely not say that, that's not Ch'ol. But you sound pretty good when you say it—you're getting much better!

Not only is establishing rapport and a clear mutual understanding good for collecting data, it is also a moral imperative. In recent decades, linguists have begun to recognize that earlier colonialist models of linguistic fieldwork—arrive in a community, interview speakers and extract language data to be published years later in arcane journals—is neither sustainable nor ethical, especially when it comes to the world's many threatened minority languages. On the other hand, active collaboration between linguists and language communities has the potential to lead to benefits on both sides. While the task of language maintenance and revitalization, along with the expertise and knowledge for how to approach it in any given community, ultimately lies with the language community itself, linguists may be able to help with everything from the creation of descriptive grammars or dictionaries, organizing and archiving material, to navigating the bureaucracy of grant-writing and reporting for language programming. Long-term collaboration with a community in turn will almost certainly provide linguists with a more complete and accurate picture of the language as it is truly spoken and used (see e.g. Hinton et al. 2018, Hinton & Hale 2001, and works cited there).

The above paragraph may suggest that linguists and language communities are distinct entities, though this is not necessarily or even ideally the case. The Mayan language family is perhaps the most impressive example of the importance and impact that native-speaker linguists can have on the health and understanding of languages (England 2007). Maya-speaking linguists have not only produced a wealth of documentary and theoretical work, but have also had an important impact on curriculum development, education, and language policy—areas that are more difficult for outsider linguists to affect. This work has led both to increased support for the languages, and a deeper understanding of their fascinating grammars.

In order to fully understand the human capacity for language it is essential that we develop a better understanding of the world's understudied languages. The scientific study of human language is relatively new, and many theories about the principles and parameters of human language were developed on the basis of better-studied languages like English and French. Though huge leaps in understanding have been made in recent decades, it is crucial that linguists continue to develop and test theories on a typologically and genetically diverse set of languages. The theory of human language must account for the fact that children can acquire *any* language with ease—Niuean and Ch'ol as well as Russian and Spanish.

This study is also urgent. Today there are around six thousand languages spoken in the world— but 96% of these languages are spoken by just 3% of the world's population (UNESCO 2003). According to some estimates, more than half of the world's currently-spoken languages will no longer have living speakers by the end of this century, unless major steps are taken to reduce current trends of language loss (Hale et al. 1992). Modern-day language loss is occurring at a rate never before seen in human history. As Ken Hale writes in the introduction to the special 1992 *Language* volume on language endangerment:

> *It is part of a much larger process of loss of cultural and intellectual diversity in which politically dominant languages and cultures simply overwhelm indigenous local languages and cultures, placing them in a condition which can only be described as embattled.* (Ken Hale, 'On endangered languages and the safeguarding of diversity')

In addition to contributing to our scientific understanding of the range and limits of possible language variation, a growing body of research has shown that the health of a community's language is a good predictor of other human health and wellness factors (e.g. Walsh 2018). For communities worldwide working to maintain and revitalize their languages, language reclamation has helped lead to a strengthened cultural identity and sense of community empowerment. Also, language revitalization and economic participation is not a zero sum game: maintaining and promoting Indigenous languages does not necessarily come at the expense of the economic opportunities generally associated with speaking the regional majority language. The benefits of multilingualism for both kids and adults have been well documented, and hold true regardless of whether the languages learned are English and Spanish, or Spanish and Ch'ol.

<p style="text-align:center">***</p>

I did eventually find my way back from the village that first summer, and have continued working on Ch'ol and other Mayan languages in the years since. While learning Ch'ol did not

alter my perception of reality the way that learning Heptapod did for Dr. Banks, some of my work on the grammar of Ch'ol has helped to shape linguistic theory. More than a decade later, for example, I think I finally have a better handle on VOS~VSO word order variation in Ch'ol (Clemens and Coon 2018). In the years since that summer I have dropped my own students off in fieldwork situations, and now that I understand the true wisdom of Haviland's simple instructions—*make some friends, learn some Ch'ol*—I haven't left them with much more than this.

That first summer in Chiapas I did make friends, lifelong ones, along with a handful of godchildren. One of my god-daughters is graduating from college during the month that I am writing this article. She is already a vocal advocate for the Ch'ol language, and the pride she has in speaking her language is a model for her peers and younger family members. She has been actively involved in Ch'ol language research with me for the past few years and is now considering graduate programs in linguistics. I have encouraged her along the way because— whatever one thinks of Linguistic Relativity, Heptapod center-embedding, and our possibilities for extraterrestrial communication—one thing *Arrival* clearly gets right is that language is a powerful tool.

References

Aissen, Judith, Nora C. England, and Roberto Zavala Maldonado, eds. 2017. *The Mayan Languages.* Routledge.

Baker, Mark C. 2001. *The Atoms of Language: The Mind's Hidden Rules of Grammar*. New York: Basic Books.

Chiang, Ted. 1999. 'The Story of Your Life.' In *Year's Best SF 4*, ed. David G. Hartwell. Harper Collins.

Chomsky, Noam. 1993. *A Minimalist Program for Linguistic Theory*. MIT occasional papers in linguistics no. 1. Cambridge, Massachusetts: Distributed by MIT Working Papers in Linguistics.

Clemens, Lauren and Jessica Coon. 2018. 'Deriving verb-initial order in Mayan.' *Language*.

Coon, Jessica, Diane Massam, and Lisa deMena Travis, eds. 2017. *The Oxford Handbook of Ergativity.* Oxford: Oxford University Press.

Dixon, R. M. W. 'Ergativity.' *Language*, 55, 1: 59–138.

England, Nora C. 2007. 'The influence of Mayan-speaking linguists on the state of Mayan linguistics.' *Linguistische Berichte* Sonderheft 14: 93-111. Special edition *Endangered Languages*, eds. Peter K. Austin and Andrew Simpson. Hamburg: Helmut Buske Verlag.

Hale, Ken, Michael Krauss, Lucille J. Watahomigie, Akira Y. Yamamoto, Colette Craig, LaVerne Masayesva Jeanne, Nora C. England. 1992. 'Endangered Languages.' *Language*, 68, 1: 1–42.

Hinton, Leanne and Kenneth Hale. eds. 2001. The green book of language revitalization in practice. San Diego, CA: Academic Press.

Hinton, Leanne, Leena Huss, and Gerald Roche. 2018. 'Introduction: Language Revitalization as a Growing Field of Study and Practice.' In *The Routledge Handbook of Language Revitalization*, eds. Leanne Hinton, Leena Huss, and Gerald Roche, xxi–xxx. New York: Routledge

Macaulay, Monica. 2004. 'Training linguistics students for the realities of fieldwork.' *Anthropological Linguistics*, 46, 2: 194–209.

Malotki, Ekkehart. 1983. *Hopi time: A linguistic analysis of temporal concepts in the Hopi language*. Berlin: Mouton.

Martin, Laura. 1986. '"Eskimo words for snow": A case study in the genesis and decay of an anthropological example,' *American Anthropologist* 88, 2: 418–423.

Massam, Diane. 2001. 'Pseudo Noun Incorporation in Niuean.' *Natural Language and Linguistic Theory* 19, 153–197.

McWhorter, John H. 2014. *The Language Hoax: Why the World Look the Same in Any Language.* Oxford University Press.

Sánchez, Liliana. 2010. *The morphology and syntax of topic and focus: Minimalist inquiries in the Quechua periphery.* Linguistik Aktuell/Linguistics Today, Amersterdam: John Benjamins.

UNESCO Ad Hoc Expert Group on Endangered Languages. 2003. 'Language Vitality and Endangerment'. Retrieved 4 June 2018.

Pullum, Geoffrey K. 1991. *The great Eskimo vocabulary hoax and other irreverent essays on the study of language*. Chicago: University of Chicago Press.

Walsh, Michael. 2018. '"Language is Like Food…": Links Between Language Revitalization and Health and Well-Being.' In *The Routledge Handbook of Language Revitalization*, eds. Leanne Hinton, Leena Huss, and Gerald Roche, 5–12. New York: Routledge.

Whorf, Benjamin Lee. 1936/2000. 'An American Indian Model of the Universe.' In J. B. Carroll, ed., *Language, thought and reality: Selected writings of Benjamin Lee Whorf*, 57–64. Cambridge, MA: MIT Press.

Whorf, Benjamin Lee. 1940/2000. 'Science and linguistics.' In J. B. Carroll, ed., *Language, thought and reality: Selected writings of Benjamin Lee Whorf*, 207–219. Cambridge, MA: MIT Press.