Preface

This Element is based on a plenary lecture to the International Pragmatics Conference in Brussels in July 2023. The reception was enthusiastic enough that I thought it might be useful to write it up. The purpose of the lecture, and thus this book, was to knock the complacency out of us by drawing attention to all the things in the field of pragmatics we still don't know or only have a feeble grasp of. It was also intended as a pep-talk for our younger colleagues, to remind them actually how practically important our subject matter is. To try to get a handle on the things we don't yet know but can at least glimpse the edges of, the Element does a quick and very superficial zip through the well-developed fields of pragmatics. The book potentially has two uses. First, and its main intended use, is to help researchers in the field find lush new pastures for study. PhD students or their supervisors might therefore find it handy as an initial thing to read. There is a danger here: the author is not a spring chicken, and so much research has accumulated in many of the domains reviewed that it is very possible that I will suggest such-and-such is under-researched and be ignorant of some rather substantial recent body of work.

A second possible use is for students new to the field of pragmatics who want such an unbuttoned, whirlwind tour of what pragmatics is all about. But here there is another danger: this survey of what we know is so brief, sketchy and loose that it may give the acolyte the mistaken impression that what we do know is self-evident, imprecise and paltry compared to what we don't know! That is best countered by having in the other hand one of the thorough textbooks now available (e.g. in chronological order, Levinson 1983; Huang 2007; Senft 2014; Clift 2016). The novice will then find that pragmatics is a well-developed field of study, with a range of quite technical and advanced nooks. With that antidote, I think this book may prove useful.

One further caveat. Like the major traditions in linguistics, this work takes language spoken in verbal interaction as the basic target. There are other channels like that involved in sign languages, and many other kinds of language use, both spoken and written. All these are worthy targets of research.[1] But I start with the prejudice that it is spoken language that kickstarted the human species and is still the predominant form of language use today. So it has a special call on our attention.

1 What Is Pragmatics and Why Does It Matter?

Pragmatics is the study of how language is used to communicate. A huge amount of thinking and research over centuries has gone into the study of language,

[1] See, for example, Hoffmann and Bublitz 2017 on social media. Historical pragmatics is an important subdiscipline that relies of course on written texts, see e.g. Jucker and Taavitsainen 2010.

mostly into how languages are structured and how history has shaped them. Until relatively recently (the last fifty years), comparatively little had been done on systematic studies of language use. This is odd: it would be curious to study the structure of a spade or a hammer without asking what it was used for and how it was shaped for those purposes. The main reason for this relative neglect of language use is presumably simply that language is so much the central human medium – the water in which we swim – that we take the usage patterns for granted. A contributing factor is undoubtedly that until the invention of practical recording techniques, 'freezing' language usage for inspection was problematic.

But despite earlier neglect, in the last half century research in pragmatics has flourished. This Element can serve as an introduction to this vibrant field, for it will sketch many of the major developments and achievements in this domain (and by 'sketch' I mean that this is a fast and loose rendition, without any of the precision that can and has been brought to bear on the subject). But its main purpose lies beyond that, namely to identify what remains mysterious and little understood about how we communicate with language, and in this way to help direct research efforts into the future.

Why is this field important? Consider the following accident. In 1990, Avianca Flight 052 approached John F. Kennedy airport, New York, low in fuel. It was put into three holding patterns until the fuel level was critical, finally missing the runway and crashing into a hillside on Long Island with the loss of seventy-three lives. There had been repeated communications between the cockpit and the control tower, with the co-pilot repeatedly mentioning that they needed priority and were running out of fuel. But the crew failed to use the word 'emergency', which is the fixed expression along with 'mayday, mayday', to request priority landing rights. Because this fixed form was not used, the control tower assumed the situation was not critical. The assumption was based on a normal rule of thumb governing language use – if someone doesn't use the extreme end of a scale, they do not intend it, so for example saying 'The crucible is still warm' suggests it is not still red hot (see Section 7). Or consider another case. Edward aged three was delayed in language abilities, and the medics diagnosed him as autistic, which put him into specialist nurseries and schools. It turned out to be a complete misdiagnosis. Edward was largely deaf. Once that was recognized, and he was fitted with hearing aids, he could fully participate in normal school.[2] Autism in fact has a very clear pragmatic profile in social interaction, with delays and disconnectedness in response and distinctive gaze patterns. Understanding the pragmatic profile is crucial and

[2] www.ndcs.org.uk/information-and-support/parenting-and-family-life/families-magazine/your-stories/primary-years-stories/edwards-misdiagnosis/

would aid early diagnosis. Still not convinced about the importance of pragmatics? Then consider this too. There are sustained efforts to make AI devices interact with us, and people with physical hindrances can come to rely on them. But these systems – Alexa, Google Assistant or the more sophisticated systems too – have no competence to cope with mishearings or incomprehension, whereas real language users have multiple systems for correcting, rephrasing and compensating. AI systems desperately need a human-oriented pragmatics.

Perhaps these illustrations of the importance of the field will seem rather marginal. Well, then, consider the case of the child learning her first language. Noises are being made around her. How should the child realize that these are communicative? Mum makes noises to the child, the child smiles back, the mother laughs, and the two are engaged from early days in an exchange. The child brings to all this some kind of knowledge or instinct about communicative interaction, and it is this presumption of meaningfulness that makes it possible for the child to learn a language. In doing so, the child uses many presumptions about the use of language – for example, that it is exchanged in turns in a kind of 'proto-conversation'. These are the pragmatic foundations for language, without which all the rich resources of the full tongues cannot be mastered.

Our knowledge of pragmatics is now extensive and based on a growing body of work that extends over fifty years. Good reviews can be found in Levinson 1983; Huang 2007; Clift 2016, to mention just a few. But the main purpose of this book is to first glance at what we know, but then look beyond that, to what we do not yet understand, and so try and discern targets for future research.

2 How to Find out What We Don't Know We Don't Know

Dark Matter makes up 85 per cent of the mass of the universe, but it is invisible to current astronomical methods. We know it must exist only because of the gravitational effects it exerts on the heavenly bodies we can see. If despite thousands of years of careful observation of the heavens, we only understand at most 15 per cent of what moves the celestial bodies, then we can be fairly confident that in the short lifetime of the scientific study of pragmatics we understand rather less than that. The rest is the 'Pragmatic Dark Matter' of my title. Much of our mental life is hidden from view: we do not know, for example, how we mentally decide what to say and how to say it. We do not know why we dream, or where bursts of inspiration come from. In the same way we don't understand many aspects of how we come to construe particular utterances in the way we do in a specific context.

There are a number of specific reasons to think there is actually a lot of pragmatic Dark Matter. First, as mentioned, the subject is scarcely half a century

old. Secondly, a great deal of the theory that we rely upon dates back to the foundational period, roughly 1960–1985, with forbears in the 1950s (Wittgenstein, Bar-Hillel, Carnap, Bühler and others). Early in that period the philosopher Austin (1962) introduced the notion of speech act (later systematized by Searle 1969); Grice (1967) introduced the notion of conversational implicature building on his earlier intentional theory of meaning (Grice 1957); Schegloff & Sacks (1973) introduced the principles of conversation analysis, Stalnaker (1974) building on earlier work by Strawson (1950) tried to firm up the notion of presupposition, and Fillmore (1971) systematized what we then knew about deixis. A great deal of further analysis and reanalysis quickly followed (e.g. Horn 1972; Levinson 1983; Sperber & Wilson 1986). These insights still form the core of our theoretical apparatus in pragmatics, but it is getting quite old and is surely ready for a refresh.

A third reason to suspect that we have hardly begun our explorations of language usage is that most of this theory derives from Western philosophy, and inevitably reflects the foci and preoccupations of Western scholars and societies, that is, the cultures of the Global North. Indigenous theory from elsewhere, and particularly from Asia where there are long traditions of metalinguistic thinking, would help to correct this viewpoint.[3] A fourth and related point is that pragmatic theory and analysis is very much focussed on familiar major languages, and indeed largely on European ones and English in particular. We have systematic information about the use of only perhaps 5 per cent of the world's languages (a good third of languages don't even have any grammatical descriptions; Skirgård et al. 2023). But there are some 7,000 languages spoken or signed on the planet, each with their own peculiarities, and information about their usage will certainly lead to new insights and fresh theory.

A fifth reason to think there is still much to discover is that much of the work done in pragmatics has been done with relatively unsophisticated tools, more akin to the astrophysics of Galileo or Newton than the modern world of space telescopes. But new tools and methods are increasingly becoming available. Digital video on personal computers only became available in the 1990s, making possible annotated video for the first time around 2000. It is only very recently that we have large multimodal corpora online and the facility to readily script search procedures, using for example machine learning to find the target phenomena. Recordings with multiple cameras, multiple sound channels, and time-aligned simultaneous recordings of eye-movement, ultrasound recordings of the vocal tract, heart-rate, breathing and other channels are now available.

[3] See for example Hanks et al. 2019.

Then there are all the resources of neuroimaging, from EEG, MEG to MRI to investigate. All these new tools and methods will throw up new phenomena we had little idea even existed.

A sixth reason to think there's a great deal still to discover is that there are still many under-developed topics of research – for example, the pragmatics of sign languages, newly described pragmatic disorders and how to improve language use of artificial agents. Interesting questions like whether pragmatic routines (like gesture repertoires, prosodic patterns, address usage) cross language-boundaries within so-called 'language areas' (e.g. the Indian subcontinent, or Meso-America) have hardly even been raised.

It is an interesting conundrum to wonder how we can convert *unknown unknowns* – that is, things we have no inklings about whatsoever – into *known unknowns*, that is, Dark Matter. In the case of astrophysics, it is by mathematically discerning the hidden forces that must account for the celestial observations. In the study of language usage one thing we can do is traverse the *known knowns*, the things we think we understand, and see just where these bump up against the edges of our known universe. That will give us a clue to what must lie just beyond our known boundaries, the known unknowns to which we should be directing our attention in the future. So, to explore these edges, this little book will take us for a wild romp through the known knowns of pragmatics in search of the known unknowns. Fasten your safety belts please!

3 The Human Communication Bottleneck and the Niche for Pragmatics

There is one necessary preliminary. Human communication is a miracle. There is nothing else like it on the planet. We are the only animals that can communicate thoughts of arbitrary complexity to each other. In this context, it might seem churlish to point to a major flaw in the design, as it were. But there is one. The fly in the ointment is a tight bottleneck on speech production. There's a physiological maximum of about seven to eight syllables per second (Laver 1994). The reasons for this are numerous. Over one hundred muscles are involved in speech production, breathing needs to be coordinated, decisions have to be made about what to say and how to say it, lexical items need retrieving, the words need to be tied together within a grammatical frame, the whole has to be phonologically encoded and finally articulated (Levelt 1989).

To get a perspective on this slow speech encoding process it is useful to convert the measure into the universal language of data transmission, namely 'bits'. Using the information theory devised by Shannon and Weaver (1949), we can calculate a maximal language data transmission speed of 96 bits per second

(bps) for English (following Laver's syllable rate above).[4] If potential transmission speed is near to 100 bps, actual trends measured across numerous languages are nearer to 39 bps (Coupé et al. 2019). Now if you compare that to your typical broadband speed of 30,000,000 bps or higher one gets an interesting perspective on the speed of human language production: it is brutally slow! Psycholinguists have also ingeniously measured exactly what part of the encoding process takes how long – the standard figure is that retrieving and saying a single word will take well over half a second before anything comes out of your mouth (Levelt 1989; see Section 10).

Such slow data transmission speeds are counter-intuitive: the phenomenology is of fast and furious conversational exchange, and certainly not of plodding effort. But these facts are firm. It is important to note though that bit rates measure data transmission (coded sequences) measured in inverse probability of occurrence, not in the amounts of semantic content transferred (Floridi 2010). Semantic information is much harder to measure, and just about the only useful measure we have is Carnap & Bar-Hillel's 1952 theory of semantic information, which shares the inverse rule (information content increases inversely to probability). A useful way to think about this is that an assertion is semantically informative to the extent that it rules out states of affairs. So 'All men are mortal' is more informative than 'Englishmen are mortal'.

Now, it is quite easy to show that semantic information can be transferred much faster than we can actually speak: if you take a tape or recording of fast speech and speed it up three times you can easily understand it, and you will likely understand it even at four times the maximal speaking rate.[5] That shows that the bottleneck is a coding bottleneck, not a limit on comprehension. This gap between the coding rate and the comprehension rate is of fundamental importance for our subject. For this is *the pragmatic niche*, the zone that can be filled by ancillary means of communication. The rest of this Element is about how pragmatics fills this gap between frustratingly slow production and fast comprehension.

Now, one response to all this may be that human communication is not really characterized by supreme efficiency – people chat, lament the weather and express their social relationships in an extravagant use of banter. Yes, but that is not the point. Language use tends to optimal efficiency even when engaged in the practice of bullshit (Frankfurt 2005). Those playing golf or tennis are complete time wasters, but they are also trying to win efficiently. Just as wanderers through a park tend to make a path by taking the shortest route between two points, so

[4] The details of the calculation can be found in Levinson (2000: 382 n. 18 & n. 19).
[5] See e.g. Laver 1994:543, Gransier et al 2023.

words tend to shorten according to how often they are used (Zipf 1949). Humans are optimizers in communication as in everywhere else.

4 A Design Perspective on Human Communication

It is sometimes very useful to adopt a design perspective when thinking about any human or animal capacity: what would it take to retro-engineer the system under study? So let us suppose we were trying to build a language with maximal expressive or communicative effectiveness. Let us appoint an engineer to help us. Obviously, our engineer would try and maximize the data transmission rate. In the Shannon–Weaver model of communication sketched in Figure 1, the effective bit rate depends on the noise in the channel and on the construction of the language.

Noise is best countered by building in a certain amount of redundancy, and languages do this in multiple ways, for example by multiple marking of such grammatical categories as plural, as in '**Women are** the best **writers**'. The bit rate can be increased by increasing the number of phonemes and the total possible syllables – recollect that data transmission in bits is inversely related to probability of occurrence, so the rarer the segment or syllable the more 'informative' in this data transmission sense. But too many phonemes and too many syllables makes a language hard to learn and slow to speak, so in practice it's best to just pick a median figure that will give us an average 40 bps speed (Coupé et al. 2019). Incidentally, one might wonder if sign languages are constrained in the same way: although manual signs are slower to produce than spoken syllables because the articulators are much larger, this may be compensated for by the two hands and facial gestures allowing simultaneous broadcast of signals (Wilbur 2009). Given that simultaneous interpretation of sign languages to spoken languages and vice versa appears to operate at near equal speeds, we can presume that sign language encoding has roughly the same bit rate as speech (see also Grosjean 1979).

So, if the bit rate is fixed, the next thing we and our engineer will worry about is making sure we can maximize semantic informativeness, and here using the

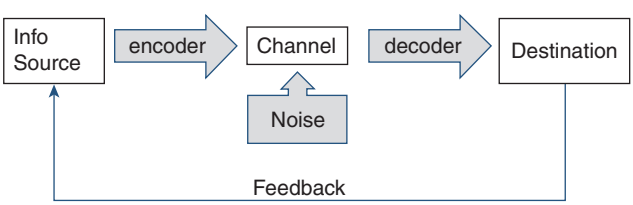

Figure 1 The Shannon–Weaver (1949) model of communication/data transfer.

Carnap and Bar-Hillel metric (informativeness in proportion to excluded possibilities) we need to ensure that the language is capable of general statements that rule out the most possible states of affairs. So we need a lot of very broad, semantically general words, like 'person', 'thing', 'tree' and so forth, and we need quantification and negation so we can say, for example 'No women are immortal either' or 'There has never been a five-legged animal'. In short, we'll need the full apparatus that natural languages have to express logical relations using a general vocabulary.

Having done his or her best to maximize the bit-rate and the expressive potential, the next thing the engineer wants to do is somehow get around that slow speech production rate. It is frustrating that, as mentioned in section 3, numerous experiments show that we can understand much faster than we can transmit. So what the engineer will try and do is devise some ways to get around that bottleneck. His or her job is to find some tricks that will amplify the content without, alas, being able to speed up the transmission.

If our engineer is any good, he or she will come up with at least five tricks – devices to circumvent that speech production bottleneck and utilize the full potential of the gap between slow speech and fast comprehension. The rest of this Element will explore these five tricks. Each of them, it turns out, is a rich domain of pragmatics, already partially explored. So, in examining them in the following sections, we will be traversing the *known knowns* of pragmatics, in search of the edges of our knowledge, the discernible *known unknowns*, the targets for future research. For each trick to circumvent the bottleneck, we'll first describe what we know, and then turn to what these aspects indicate that we do not yet fully understand.

5 The First Trick to Circumvent the Bottleneck: Multiplying Channels

5.1 The Known Knowns: Multimodality

The first trick is a no-brainer. We have a strict coding speed limit on the speech channel. Very well, we'll use other channels as ancillary devices, for example gesture. All natural languages used in social interaction use multiple channels. For example, I can say 'He went that way' indicating leftwards with my hand. Or I can say 'The boss says redo it' while rolling my eyes, indicating disaffiliation with the message. Sometimes these different channels work to give the desirable redundancy (I say 'He turned left' while gesturing left), but more often they add new information. How many distinct channels or conduits of information are there actually?

Here we should pause to distinguish *the channel*, for example vocal-auditory vs. gestural-visual, from the *medium*, for example English vs. Swahili. English

can be delivered in the visual channel by writing of course, or it can be finger spelled in sign language, hence the need for the distinction (Lyons 1977: Chapter 3). But that won't be sufficient. We also need a notion of layering. Layering of one sort is a notion familiar to linguists through for example the distinct levels of phonetics and phonology, where elements of one layer constitute elements of another. But here we need a notion of *overlayering*, that is, the possibility of a layer carrying an independent (non-constitutive) signal. A contrastive prosody can be overlaid onto the segmental signal to indicate both grammatical (e.g. interrogative) functions and attitudinal ones (Ladd 2014). Part of this has been described as *paralanguage*, for example the use of creaky voice, the relaxed vocal chords associated with self-confidence. Voice quality, amplitude, pitch and timbre can all be used in this way to add essential information to an utterance.

In a similar way we can view the gestural-visual channel as layered. I can gaze at you with narrowed or wide-open eyes, momentarily or in a sustained manner, with blinks or without blinks. A hand gesture to the left can be close to the body or extended, made with a loose hand or an index finger, and so forth. Each layer carries potentially independent meaning. As Abercrombie (1968: 55) put it, 'We speak with our vocal organs, but we converse with our entire bodies'.

So here is a way that our engineer can get around the coding bottleneck, by multiplying channels and within channels by multiplying layers. Given the number of potential channels and layers (some sketched in Figure 2) one can see that this can easily multiply the bit rate, the transmission speed, of human communication. Ray Birdwhistell, who invented the term kinesics for the study of bodily posture, estimated that at most a third of the content of human communication is verbalized.[6] That measure is suspect, but there is no doubt that the use of multiple channels can amplify and sometimes multiply the content of the spoken channel.

The study of all this is not new. The Greek and Roman orators already classified gestures, but systematic study began in the second half of the twentieth century, on the one hand by anthropologists interested in gesture and bodily deportment and on the other hand by phoneticians interested in the layering of the verbal signal. In addition, social psychologists got increasingly interested in non-verbal communication. A long-running project on the '*Natural history of an interview'* involved many of the crucial pioneers in the non-vocal channels, including Birdwhistell, Hall, Kendon, Condon, Erikson, Bateson and others,

[6] McDermott, R. 1980. Profile: Ray L. Birdwhistell. *The Kinesis Report*, 2(3): 1–16 (cited in Wikipedia article 'Kinesics', https://en.wikipedia.org/wiki/Kinesics#cite_note-7)

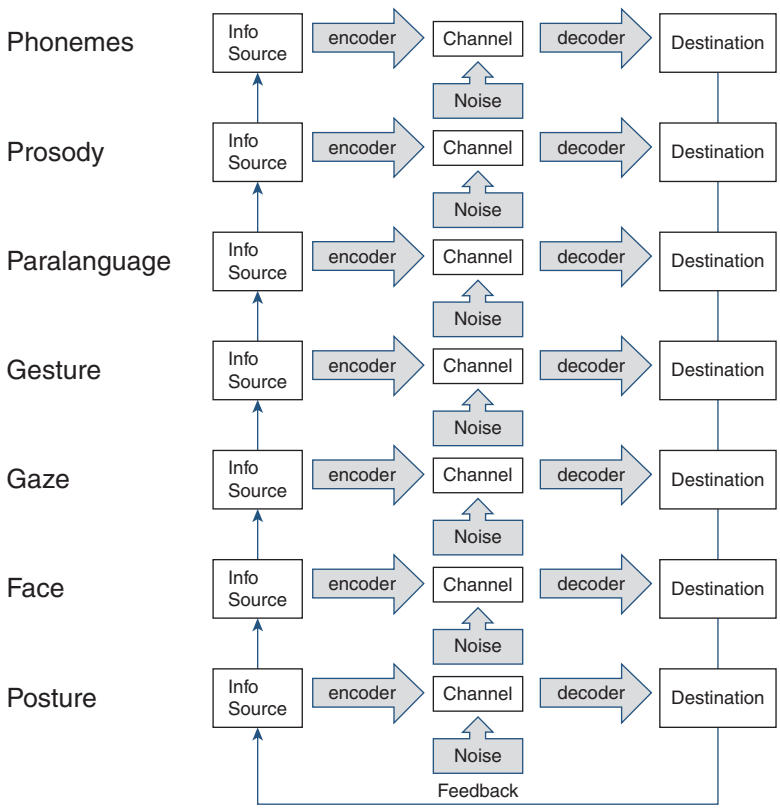

Figure 2 Circumventing the coding bottleneck by multiplying channels and layers. *The figure is only illustrative – how many channels are there actually?*

who developed the first ways of describing and annotating filmed interaction.[7] These early scholars were mavericks outside of mainstream disciplines, but with videotape and then digitized video, and from about 2000 with the birth of digital video annotators like ELAN,[8] the study of the multi-layered nature of human communication has developed very rapidly. Recent technical progress is beginning to make it possible, using machine learning, to automatically retrieve facial expressions and particular gestures.

Pragmatics has embraced all these developments, under the rubric of 'multi-modality', and the study of manual gesture in particular is well advanced. There are standard ways of breaking down hand movements, for example, into preparation phases, the stroke or main gesture, its potential hold, and then retraction (Kita et al. 1998). Interestingly, gestures seem to be integrated into

[7] See https://www.lib.uchicago.edu/mca/mca-15-098.pdf [8] https://archive.mpi.nl/tla/elan