



What speakers do

- Introduction
- Speech production
- Conceptualization and formulation
- Articulation
- Self-monitoring and repair
- Automaticity
- Fluency
- Managing talk

Introduction

'My students always say that they want more speaking, but I don't know how to teach it, apart from giving them lots of useful expressions.'

'I've been asked to teach a conversation class, but what is conversation? Is it just free speaking?'

'How much grammar do students need before they can have conversations?'

'How can I help my students become more fluent? What is fluency? Is it good pronunciation?'

'My business students are good at giving presentations, but they can't have even the simplest conversations. How can I help them improve?'

Questions like these – from a teachers' on-line discussion forum – may be familiar to you. They express some of the common dilemmas teachers face when trying to address the teaching of speaking. For a long time it was assumed that the ability to speak fluently followed naturally from the teaching of grammar and vocabulary, with a bit of pronunciation thrown in. We now know that speaking is much more complex than this and that it involves both a command of certain skills and several different types of knowledge. In this chapter we will look at speaking from the first of these perspectives: what is it that good speakers *can do*? In the chapter that follows we will address the second question: what is it that good speakers know?

Speech production

Speaking is so much a part of daily life that we take it for granted. The average person produces tens of thousands of words a day, although some people – like auctioneers or politicians – may produce even more than that. So natural and integral is speaking that we forget how we once struggled to achieve this ability – until, that is, we have to learn how to do it all over again in a foreign language.

What then is involved in speaking? The first point to emphasize is that speech production takes place in real time and is therefore essentially **linear**. Words follow words, and phrases follow phrases. Likewise, at the level of **utterance** (that is to say, the spoken equivalent of sentences), speech is produced utterance-by-utterance, in response to the word-by-word and utterance-by-utterance productions of the person we are talking to (our **interlocutor**). This **contingent** nature of speech, whereby each utterance is dependent on a preceding one, accounts for its **spontaneity**. This is not to say that speech is unplanned, only that the planning time is severely limited. And the planning of one utterance may overlap with the production of the previous one. These 'real-time processing' demands of speech production explain many of the characteristics of spoken language.

In the following extract from a dinner party conversation about traditional British foods (which we will call *Kedgerree*, after the fish dish it names, for ease of reference) we can get a sense of speech production in operation. There are five speakers, and the subject of *junket* (an English milky dessert) has come up. One of the speakers, Kath, then says:

(1) Kath: I made junket when I was in the first year of secondary school.

(The numeral (1) is the way speaker **turns** are usually indicated in transcriptions. A **turn** is the duration of one speaker's contribution to the talk before yielding to, or being interrupted by, another speaker). Other transcription conventions are:

- = contiguous utterances, i.e. ones that run on without pause, despite interruptions from other speakers
- | overlapping utterances
- || simultaneous utterances
- () a slip

The conversation about junket meanders on, until 51 turns later Kath says:

- (52) Kath: It's one of those ridiculously old-fashioned dishes that they make you cook in domestic science =
- (53) Hilda: This is really nice this Rioja
- (54) Nick: Well why don't you try making | some? Might be great
- (55) Kath: | = like kedgerree
- (56) Simon: Spotted dick.
- (57) Kath: = Kedgerree, I remember saying to my mum =
- (58) Scott: Toad-in-the-hole
- (59) Kath: = I've got to take a pound of fish next week we're making kedgerree and she said [mock accent] 'you don't want to be making kedgerree' [laughter] and she said 'we don't like it'. And I had to take a note to my domestic science (taitch-) teacher saying 'Kathleen can't make kedgerree because we don't like it'. [laughter] Awful. So I couldn't make it. I had to sit there while everybody else did. [laughter].

Continues ...

- (60) Hilda: I would just make egg and bacon
 (61) Kath: But kedgerree. This was a sort of comprehensive school the first year of. Nobody knew what kedgerree was. It was sort of kedgerree and junket [laughter]
 (62) Simon: || I love kedgerree
 (63) Kath: || I mean for God's sake
 (64) Simon: Have you ever eaten kedgerree since?
 (65) Kath: || Oh yes I love kedgerree
 (66) Nick: || Didn't you say you could get hold of a decent bloody
 (67) Kath: It's a sort of old colonial dish =
 (68) Simon: = It is yes it's Indian
 (69) Kath: = like junket is but it was so | inappropriate =
 (70) Nick: | oh is it like galub jalum?
 (71) Kath: = for the first year comprehensive school kids to be making [laughs].

Conceptualization and formulation

The mention of *junket* seems to have triggered an association in Kath's mind with domestic science classes (turn 1), which in turn reminds her of a story about – not junket – but kedgerree. At some point (it may have been at the initial mention of *junket*) she **conceptualizes** the story – in terms of its discourse type (*a story*), its topic (*kedgerree*), and its purpose (*to amuse*). She then has to wait for the appropriate moment to re-introduce it (turns 52 and 55), where she adroitly shifts the topic from *junket* to *kedgerree*. Finally, at turn 57, she is able to 'gain the floor' and is ready to tell her story.

But first the story-idea has to be mapped out, or **formulated**. This involves making strategic choices at the level of discourse, syntax, and vocabulary. At the level of overall discourse, stories have a typical structure, or **script**. At the very least, they have a beginning, middle, and end. Discourse scripts are part of our shared background knowledge, and can be 'pulled down off the shelf', as it were, thereby saving formulation time, while also easing the load of the listener, who quickly recognizes what script has been selected.

Each of the stages of the script then needs to be fleshed out at the utterance level. This is where the specific **syntax** of each utterance needs to be chosen so that the content of the story is packaged in a way that is consistent with the speaker's intentions. Initially, this will mean deciding on what elements of the utterance will go in what order. In English, utterances tend to have a two-part structure: the first part is the **topic**, i.e. what we are talking about, and the second part is the **comment**, i.e. what we want to say about the topic. So, in Kath's turn 67: *It's a sort of old colonial dish*, the topic is *it* (referring to the previously mentioned *kedgerree*) and the comment is everything that follows:

topic	comment
It [kedgerree]	is a sort of old colonial dish.

The topic is typically information that has already been mentioned (**given information**), while the comment is usually something new.

The ‘grammaring’ of each utterance is also constrained by how much information can be held in working memory at any one time. One way speakers compensate for limited planning time is to use what is called an **add-on strategy**. This is the chaining together of short phrases and clause-like chunks, which accumulate to form an extended turn. We can see the add-on strategy operating in Kath’s turn 61 in the Kedgeree conversation. Each stage in the sequence is marked with a vertical line:

But kedgeree.| This was a sort of comprehensive school | the first year of.|
 Nobody knew what kedgeree was.| It was sort of kedgeree and junket

If this had been a written sentence, it would probably have been constructed quite differently, with more embedding (or **subordination**) of components, rather than simply chaining them together. (See page 53 for an example of how this might be done.) This accounts for the often fragmented appearance of spoken language when it is transcribed. In listening to spoken language we tend to ‘iron out’ its creases, so that we hear it as a smooth continuum.

Having ‘laid out’ the utterance in terms of its syntactic elements, the speaker now needs to assign individual words or phrases to the different ‘slots’ in the layout. Take, for example, Kath’s decision (in turn 57) to use the words *my mum* to fill the slot at the end of *I remember saying to ...*. Kath’s choice of the word *mum* rather than *mother*, for example, may be due to the fact that she uses this term more frequently when talking about this person. On the other hand, the choice may be determined by an assessment of how appropriate the word is for the particular context. For the purposes of the anecdote that Kath is telling, which pokes gentle fun at her mother, the more informal *mum* probably felt more appropriate.

When a wrong word is accessed, it will usually be a word – or combination of words – that is similar in meaning or form, or both, to the targeted word. Production slips often occur when speakers are under pressure or tired. Sports commentators, for example, are particularly susceptible to this, and collections of sporting gaffes circulate freely on the Internet. For example:

‘I can see the carrot at the end of the tunnel.’ (Stuart Pearce)
 ‘The tackles are coming in thick and thin.’ (Alan Brazil, on Radio 5 Live)
 ‘Barnsley have started off the way they mean to begin.’ (Chris Kamara)

At some stage, the words need to be ‘glued together’ by the insertion of the appropriate grammatical markers, such as articles (*the, a* etc), auxiliary verbs (*is, did, have* etc) and word endings (*-ing, -er, -ed*, and so on). Occasionally, even native speakers make slips at this level, producing *I buyed it* for *I bought it*, for example, although they are always quick to correct these slips.

Also at the formulation stage, the words need to be assigned their pronunciation. This will include not only the individual sounds of the words but the appropriate placement of prominence (**stress**) and the meaningful use of **intonation** (pitch direction). Because sentence stress and intonation are implicated in the way new or important information is signalled, it is likely that choices at this level have been made at the initial ‘laying out’

stage. So, when Simon asks (in turn 64) *Have you ever eaten kedgerree since?*, the key piece of new information is not *kedgerree*, which has already been mentioned, but the last word, *since*. This is therefore given extra prominence. Also, because he's asking a *yes/no* question, the pitch direction rises on the word *since*. In Kath's reply (turn 65), prominence is given to the word *love*: *Oh yes I LOVE kedgerree*, and the direction of pitch change is down, conveying the completion of the idea posed by Simon's question.

Articulation

What has been formulated now needs to be articulated. Articulation involves the use of the organs of speech to produce sounds. A stream of air is produced in the lungs, driven through the vocal cords, and 'shaped' by, among other things, the position and movement of the tongue, teeth, and lips. Vowel sounds are produced primarily by the action of the tongue and the lips. Consonant sounds are determined by the point at which the air stream is obstructed – e.g. at the lips or teeth – and the kind of constriction the air stream is subjected to, e.g. whether it is made to 'pop' or to 'buzz'.

The combined effect of all these variables allows speakers of English to produce a range of over 40 **phonemes**, i.e. sounds that, in English, determine the meaning of a word. These are divided almost equally between vowels and consonants: the exact number will vary according to the variety of English spoken. (For a fuller description, see *How to Teach Pronunciation* by Gerald Kelly.)

The rather oversimplified account, above, might suggest that individual sounds are produced one at a time, in the manner of a sausage-machine. In fact, sounds are produced in a continuous stream, with many different vocal organs involved concurrently, such that the articulation of one sound will affect the articulation of its neighbours. This accounts for the way, in fluent speech, some sounds merge with other sounds (as in the way *handbag* sounds like *hambag*) or are dropped completely, as is the final *d* in *baked beans*.

At the same time as these articulatory processes are engaged, continual changes in loudness, pitch direction, tempo, and pausing serve to organize the sounds into meaningful word forms, and the words into meaningful utterances. All this physical work happens, of course, at great speed. It is estimated that proficient speakers produce 15 phonemes a second. Sometimes, in the rush to speak, interference from neighbouring words causes pronunciation slips, as when Kath says (turn 59): *And I had to take a note to my domestic science taitch- teacher saying ...* Here, her anticipation of the /eɪ/ sound in *saying* seems to have interfered with the correct pronunciation of *teacher*, which she starts to pronounce as *taitcher*.

Self-monitoring and repair

This quick self-correction of Kath's is an instance of **self-monitoring**, a process that happens concurrently with the stages of conceptualization, formulation, and articulation. A re-think at the planning stage may result in the abandonment of the message altogether, as when someone starts to gossip and then realizes that the subject of the gossip is within hearing distance! Self-monitoring at the formulation stage may result in a slowing down, or a pause and the subsequent backtracking and re-phrasing of an utterance. Self-monitoring of articulation results in the kind of corrections

that even fluent speakers have to make when the wrong word pops out or the pronunciation goes awry.

Hand in hand with monitoring is the ability to make running **repairs**, either in response to self-monitoring or to the messages conveyed by one's interlocutors. Repair can take the form of an immediate correction (as in Kath's *taitch- to teacher*) or 'retrace-and-repair' sequences, that is, when the speaker retraces or 're-winds' an utterance, and starts again, but with a different sequence of words or phrases, as in:

Dad, I don't think you sh-, I think you should leave Chris home Saturday.

Automaticity

All this conceptualizing, formulating, articulating, and monitoring mean that a speaker's attentional resources are very thinly stretched. In order to achieve any degree of fluency, therefore, some degree of **automaticity** is necessary. Automaticity allows speakers to focus their attention on the aspect of the speaking task that immediately requires it, whether it is planning or articulation.

At the level of formulation, automaticity is partly achieved through the use of **prefabricated chunks**. Kath saves valuable planning time by using expressions like *I remember + -ing*, as in *I remember saying to my mum ...*. This is an instance of a chunk that acts as a kind of springboard into the anecdote that follows, one that is commonly used to introduce anecdotes. The repeated linking device *and she said* is another instance of a ready-made unit that is common in storytelling.

At the discourse level, a degree of automaticity is possible too. Kath's story is told with little hesitation and no false starts probably because she has told this story a number of times before. Not only is the overall design of the story familiar to her, but whole segments of it (such as *'Kathleen can't make kedgerie because we don't like it'*) may have been memorized from previous tellings.

In this sense, speaking is like any other skill, such as driving or playing a musical instrument: the more practice you get, the more likely it is you will be able to chunk small units into larger ones. With fewer units to assemble at the moment of articulation, there is a proportionally greater gain in fluency. Conversely, embarking on a completely untried speech genre, on an unfamiliar topic, with unknown interlocutors, is likely to make one 'tongue-tied' at best, or, at worst, completely mute. Wilbur Wright, writing of his first ventures in aeronautics, said, 'Skill comes by the constant repetition of familiar feats rather than by a few over-bold attempts at feats for which the performer is yet poorly prepared.' The same could be said for spoken fluency.

Fluency

What exactly is **fluency**, then? When we talk about someone as being 'a fluent speaker of French' or when we say 'she speaks the language fluently', what do we mean? Is fluency simply the ability to speak fast? Speed is a factor, but it is by no means the only – or even the most important – one. Research into listeners' perceptions of a speaker's fluency suggests that **pausing** is equally

important. All speakers pause – they have to, to draw breath. And even proficient speakers need to stop from time to time to allow the formulation of an utterance to catch up with its conceptualization. But frequent pausing is a sure sign of a struggling speaker. If the speaker – produces – one – word – at – a – time – like – this – no matter how accurate the results are, the speaker will not normally be judged a fluent speaker. In fact, in terms of how listeners rate a speaker's fluency, the frequency of pausing is more significant than the length of the pauses.

Also important is the appropriate **placement** of pauses. Natural-sounding pauses | are those that occur at the intersection of clauses, | or after groups of words that form a meaningful unit. | (The vertical lines in the last sentence mark where natural pauses might occur if the sentence were being spoken.) Unnatural | pauses, on the | other hand, occur | midway between related groups of | words.

Another significant factor in the perception of fluency is the length of run, i.e. the number of syllables between pauses. The longer the runs, the more fluent the speaker sounds. Studies of speakers who are 'abnormally fluent', such as auctioneers and horse-race commentators, show that such speakers hardly ever hesitate or backtrack, and take only minimal pauses for breath. Moreover, the runs between these pauses are enormous. Here is an example of a single run in a New Zealand race-caller's live commentary:

They're off and racing now and one of the best out was Speedy Cheval coming out at number two from El Red and also Florlis Fella's away fairly well, a little wider on the track the favourite Race Ruler, Twilight Time is in behind those.

Researchers found that the race-caller's fluency was due in large part to the fact that, rather than constructing each utterance from scratch, he was using prefabricated chunks – sequences he had memorized through constant practice.

Race-callers and auctioneers are, of course, somewhat exceptional. Nevertheless, we all know someone of whom it is said 'you can't get a word in edgeways', just as we have all experienced the sensation ourselves of being totally 'tongue-tied'. Somewhere in between is the notion of normal fluency.

In order to give at least the illusion of fluency and to compensate for the attentional demands involved in speech production, speakers use a number of 'tricks' – or **production strategies**. One of them is the ability to disguise pauses by filling them. The most common **pause fillers** are *uh* and *um* (also spelt *er* and *erm*, respectively). Some **vagueness** expressions, like *sort of* and *I mean* are also used to fill pauses. Another common device for gaining formulation time is the use of **repeats** – that is the repetition of a single word at the point where formulation has been temporarily paused. In this short extract, the speaker uses both fillers and repeats (the dot indicates a short pause):

well what's the · what's the failure with the football I mean this · this I don't really see I mean it · cos the money · how much does it cost to get in down the road now?

The features of fluency can now be summarized:

- pauses may be long but not frequent.
- pauses are usually filled.
- pauses occur at meaningful transition points.
- there are long runs of syllables and words between pauses.

Managing talk Interaction

So far we have been describing what speakers do as if they were speaking in a kind of vacuum, but of course most speaking takes the form of face-to-face dialogue and therefore involves interaction. Even in monologic speaking, such as lectures, political speeches, and stand-up comedy, most speakers adjust their delivery to take into account the response of their audience. In the Kedgeree conversation, for example, we saw how Kath had to work hard, conversationally speaking, both to gain the floor and to switch the topic so that she could tell her story:

(52) Kath: It's one of those ridiculously old-fashioned dishes that they make you cook in domestic science =
(53) Hilda: This is really nice this Rioja
(54) Nick: Well why don't you try making | some? Might be great
(55) Kath: | = like kedgeree
(56) Simon: Spotted dick.
(57) Kath: = Kedgeree, I remember saying to my mum =
(58) Scott: Toad-in-the-hole
(59) Kath: = I've got to take a pound of fish next week we're making kedgeree ...

Notice how other speakers are jockeying for conversational turns, introducing new topics and engaging in word play (naming old-fashioned dishes with comical names) and that this causes a certain amount of overlapping (i.e. more than one speaker speaking at once) and lack of coherence. But once Kath has wrested the topic and the floor, the others sit back and listen, recognizing that she has embarked on an anecdote.

Turn-taking

This delicate moment is successfully negotiated because the speakers are familiar with the rules and skills of **turn-taking**. The fundamental rule of turn-taking is:

- speakers should take turns to hold the floor.

This implies that no two speakers should be speaking at once, at least not for any sustained period of time. There are two further rules, although the first of these is arguably culturally specific:

- long silences are to be avoided.
- listen when other speakers are speaking.

The skills by means of which these rules are observed include:

- recognizing the appropriate moment to get a turn.
- signalling the fact that you want to speak.
- holding the floor while you have your turn.
- recognizing when other speakers are signalling their wish to speak.
- yielding the turn.
- signalling the fact that you are listening.

In some contexts, such as in a business meeting or a classroom, these interactional moves are formally signalled by, for example, the raising of one's hand. In casual talk between friends, however, it is largely the use of **discourse markers** that signals a speaker's conversational intentions. A discourse marker is to speaking what a car's indicator lights are to driving: it lets other speakers know what your intentions are. Typical discourse markers for managing turn-taking include:

- that reminds me* (= I'm continuing the same topic)
- by the way* (= I'm indicating a topic change)
- well anyway* (= I'm returning to the topic)
- like I say* (= I'm repeating what I said before)
- yes, but* (= I'm indicating a difference of opinion)
- yes no I know* (= I'm indicating agreement with a negative idea)
- uh-huh* (= I'm listening)

Strictly speaking, *uh-huh* is not a discourse marker but a **backchannel device**, i.e. a means of signalling to your interlocutor that you are listening, and, in the case of *really?* and *no!*, not only listening, but interested, shocked etc.

Paralinguistics

Negotiation of speaking turns does not rely on words alone. A sharp intake of breath and a raising of the shoulders, for example, signal the wish to take a turn. At the same time, the speaker-to-be typically glances away from the current speaker: it's not customary to start talking when looking directly at your interlocutor. During a speaking turn, little head nods from listeners tend to encourage speakers to speak faster, but if someone stands with their arms crossed, most speakers slow down. As speakers approach the end of their turn, there is a tendency to let the shoulders fall, and to re-direct their gaze back at their interlocutors, as if to say 'I've done'. The interactional use of eye gaze and gesture are known as **paralinguistics**. Of course, these paralinguistic signals apply only in face-to-face conversation. When speaking on the phone, listeners have to rely on other means, such as intonation, tempo, and pausing, to gauge the current state of their interlocutor's turn.

Conclusions

In this opening chapter we've tried to get 'inside the head' of a language speaker. What exactly happens when thoughts become utterances? We noted that speaking consists of at least three stages:

- conceptualization
- formulation
- articulation

during which the speaker is also engaged in:

- self-monitoring

Speakers achieve fluency because these processes are, to some extent, automated. The use of production strategies, such as the filling of pauses, also contributes to fluency. At the same time as they are speaking, speakers are also having to gauge the effect they are having on their interlocutors, as well as to take into account the contributions other speakers are making to the talk, both linguistic and paralinguistic. This involves an ability to manage turn-taking.

Looking ahead

So far, we have looked at the mental and physiological processes implicated in speaking. But what does a speaker need to *know*, in order that these processes are optimally realized? That is the subject of the next chapter.