

TOWARDS A THEORY OF FLUENCY

Bruce Duncan MacQueen^{1,2}

¹ Department of Comparative Literature, University of Silesia, Katowice, Poland

² Institute of Psychology, University of Gdańsk, Gdańsk-Oliwa, Poland

Key words: SLA, aphasia, microgenetic theory

*“Speak the speech, I pray you, as I pronounc’d it to you,
trippingly on the tongue...”*

William Shakespeare, Hamlet, Act III, Scene 2, line 1

SUMMARY

In applied linguistics, the term “fluency” occurs most often in the context of second language acquisition (SLA), where it serves as a generally accepted outcome measure. It pertains to speech, and more specifically, to the rate of speech production in the target language (L2), but it is not reducible to time. To speak a second language “fluently” is to speak as quickly as a native speaker, or nearly so, while maintaining an acceptable level of comprehensibility. In neurolinguistics and neuropsychology, on the other hand, fluency (measured in mostly quantitative terms) is a clinical parameter that distinguishes certain types of aphasia (fluent and non-fluent), and is increasingly recognized as a differentiating parameter in various types of dementia. The aim of the present paper is to suggest that despite the obvious differences in the meaning of the word “fluency” in applied linguistics and neurolinguistics, there is an important underlying, unifying concept. The central issue is time, or more specifically, how the brain organizes time and how this organization is reflected in behavior – in this case, verbal behavior. The properties of fluency are, as the etymology suggests, analogous to those of a fluid, especially movement and the relative absence of perceptible points and boundaries.

INTRODUCTION

Consider the following two sentences:

Smith is fluent in Spanish. (1)

The patient presents with a fluent aphasia. (2)

The same adjective, “fluent,” occurs in both (1) and (2), but this hardly seems more than a coincidence, and could even be dismissed as a homonym of sorts, were it not for the fact that in both sentences fluency is being used in the context of speech. What characteristic of speech could this be, however, that would justify using the same adjective in both (1) and (2)? On the face of it, the two sentences have little in common beyond the reference to speech. Sentence (1) represents perhaps the most common use of the term “fluency” in ordinary language, i.e. in the context of second language acquisition (SLA), and would easily be understood by the average speaker of English. Sentence (2), on the other hand, is clearly clinical jargon: “patient” as a noun is a giveaway, only physicians and other clinicians use the verb phrase “present with” in this way, and the term “aphasia” is a medical term whose meaning is not as widely known as it should be, given the frequency with which speech disturbances occur in persons with various kinds of brain damage.

The way in which “fluent” is predicated also differs significantly in the two sentences. In (1), “fluent” is directly predicated of Smith, in such a way as to indicate that as a speaker of Spanish she possesses the quality of fluency. In (2), fluency is a quality of aphasia, something that distinguishes a particular pathology of speech, “fluent aphasia,” from other pathologies, e.g. “non-fluent aphasia.” Note, however, (1) could be rewritten as

Smith speaks Spanish fluently. (3)

or

Smith speaks fluent Spanish. (4)

or

Smith’s Spanish is fluent. (5)

In (3), then, fluency appears as an adverb modifying the verb “speaks,” while in (4) and (5) the quality of being “fluent” is being ascribed, not to Smith, but to her speech in Spanish. Sentence (2) could also be re-written as

The patient is a fluent aphasic. (6)

Sentence (6) is inelegant, to be sure, but that is because it is now considered bad style to refer to a patient with aphasia as “an aphasic.” Nevertheless, (6) is a well-formed sentence, and any clinician familiar with aphasia will understand it. In this case, then, the shift from (2) to (6) is the reverse of the shift from (1) to (3), (4), or (5), i.e. in (6) fluency is predicated of the patient (though indirectly, since “fluent” modifies the predicate nominative, “aphasic”), whereas in (2) the aphasia is said to be “fluent.”

Contrary to that first impression, then, the common term “fluent” may not after all be entirely unmotivated, since the adjective “fluent” in both sentences

(1) and (2) seems to refer to a particular quality of speech (or of a speaker *qua* speaker, a person who speaks in a certain way). The point is hard to force, however, since multiplicity of meanings is hardly a rare phenomenon in natural languages. For example, such common words as “tense,” “mood,” and “case” often bear meanings in grammatical and linguistic terminology that have only a slight historical relation (or none at all, in the case of “tense”) to the way these same words are used in ordinary speech. Could the same not be true of “fluent” and “fluency”, that usage in different fields of thought and work has given these words different meanings in different contexts? Falling back on comparison to grammatical vocabulary does not really resolve the issue with fluency, however, since in both (1) and (2) the reference is linguistic, i.e. within the same (or nearly the same) domain and not across domains. We are entitled to wonder, at least, if there may not be after all some common ground. In the present paper, then, I will attempt to demonstrate that there is a significant link, a point of junction, and further, that the nature of this link or point is not either historical or trivial, but rather reveals some important aspects of how speech and language actually work.

The word “fluency,” like many other English words and the great majority of abstractions, is derived from Latin. To be somewhat more precise, the intransitive verb *fluo* ‘to flow’ gives a present active participle *fluens* “flowing,” from that an abstract noun (rare in Latin) *fluentia*, anglicized by normal derivation rules to “fluent” and “fluency” respectively. From the same stem the Latin adjective *fluidus* is formed, giving us both a noun and an adjective in English, viz. “fluid.” Fluent speech, then, would seem to have at least some of the characteristics of a fluid, such as motion and the absence (at least on the macroscopic scale) of parts or particles. A fluid moves in a characteristic way which makes it nearly impossible to distinguish the movement of some portion of the fluid from the motion of the whole. Interrupting the motion of a fluid, then, causes a characteristic disruption of motion (and shape) which extends over a much larger area than the source or point of the disruption itself. A fluid is not shapeless and has boundaries, but the shape is at once both remarkably persistent and remarkably changeable, as in the shape of a river or the sea, or even the water in a fountain, in which every particle is in constant motion, and yet the whole fountain gives the impression of being a stable object, something that is *here* and not *there*.

It would seem reasonable, then, to bear these and related physical properties of a fluid in mind as we consider what fluency means in the context of SLA, and what it means in neurolinguistics.

FLUENCY IN SECOND LANGUAGE ACQUISITION

Fluency in the target language is the obvious goal of foreign language instruction, the measure of success and (obviously) a key element in adver-

tising. Indeed, the concept is so closely bound up with SLA that it seems distinctly odd, even perverse and misleading, to speak of being fluent in one's native language. Even if the importance of fluency in the target language seems beyond question, however, it is a trickier problem to define what fluency is and establish how it can be measured (Douglas 2001, Norris & Ortega 2000, 2003, Chambers 1998). Despite many efforts to develop tests that would unerringly separate the fluent from the non-fluent users of a second language, assessment always seems to come down to a largely subjective judgment: one simply knows that a given speaker is fluent in the second language (L2), without necessarily being able to point out the features of speech that would justify such an impression. Errorless speech is neither a sufficient nor necessary condition, though there is surely an upper limit to the number or nature of allowable grammatical or lexical errors. Speed is likewise an important element of fluency, perhaps necessary but not sufficient, and the same can be said for comprehensibility. Fluent speech, as the name implies, *flows*, or rather, *flows smoothly*. If the flow of words is too slow or too turbulent, too often interrupted to search for words or self-correct errors in grammar or diction, then fluency is disrupted.

As a general rule we speak of the fluency of the speaker; it seems odd or wrong, again, to speak of "fluent listening" or even "fluent writing," though the point is at least arguable. We make a fluency judgment, whether consciously or not, on the basis of the oral utterance. Moreover, even the impeccable pronunciation of a very short text, i.e. the answer to a question or a simple formula of greeting, is not a large enough sample to make a fluency judgment. Nor is it sufficient for the speaker to produce a memorized or stereotyped text: we need spontaneous speech, preferably in conversation, in order to determine whether or not the speaker is fluent in the language being used. Good actors can memorize and deliver lines in a language they do not speak (as for example Kevin Costner speaking Sioux in the film *Dancing with wolves*), without the need to understand any of the words they are actually using. Yet good comprehension in L2 is not sufficient. We do not speak of fluency in the classical languages, for example, even though over a lifetime one can learn all the rules of grammar and most of the vocabulary of Greek and Latin, so as to read the texts of classical authors quickly, easily, and with full comprehension. Yet only a handful of enthusiasts ever really learn to speak the classical languages, which remain for us (ironically enough, given the importance of "oral culture" in classical antiquity) written languages only. No small difficulty is posed by the fact that we cannot be sure exactly how either language was pronounced at a given place and time, nor do we have any access to the nature of simple, ordinary, spontaneous conversation. Thus it is hardly surprising that the topic of fluency seldom arises in the pedagogy of Greek and Latin, though in recent years a pedagogical approach called "Fluent Latin" has made an appearance (see for example <http://www.campanian.org/education.html>). The distinction between mastery of ancient and modern Greek, both in terms of method and results, is instructive in this respect.

The synthesized speech of computers also notoriously lacks fluency, which partly derives from the mechanics of synthesizing phonemes, partly from the inability of all but the most powerful and sophisticated computers to follow the heuristics of conversation. Synthesized speech is based on the assumption that a verbal utterance is built in a manner analogous to the way letters are combined to make written words, which in turn is the way the binary language of computers works. Thus a spoken word is logically assumed to be a string of spoken phonemes, so that to pronounce a word it should be enough to determine what phonemes go into its construction and then realize them in the appropriate order. It would be a serious mistake, of course, to make the naive assumption that this can be done by converting graphemes into phonemes directly, especially in English – how would a computer thus programmed say the word *knight*? – but this problem can be avoided by converting the word into phonemes according to a standard algorithm, such as the IPA. Thus the troublesome *knight* becomes /najt/, composed of the three phonemes [n], [aj] and [t], which can then be realized acoustically by the computer. The problem is – as students of applied linguistics hardly need to be told – that this approach does not produce fluent speech in any language. A phoneme is not strictly defined acoustically; it is, rather, a range of possible acoustic realizations, i.e. allophones, regarded by native speakers of a given language as being “the same.” The choice of allophones at a given moment in time is conditioned by many factors, both individual and general: the person’s accent, state of mind, etc., the phonemes immediately preceding or following, the accentual structure of the word in which the phoneme occurs, and so on. Thus there exists a veritable host of so-called suprasegmental aspects of speech to be considered, each of which can motivate the choice of an allophone – or not. All this uncertainty and the multiplicity of factors taxes the computing power of even the most advanced computers.

Fluency in a foreign language does not require, of course, that the accent be perfect. The phoneme map of the non-native speaker almost never coincides exactly with that of the native speaker, since the constant use of a given language (or languages) in childhood sculpts the muscles and ligaments of the speech organs in a particular way. Nevertheless, a positive fluency judgment assumes that the allophones used, even if not precisely those that would be used by a native speaker of L2, do not cross phoneme boundaries, so that the phonological realization of the individual words does not disrupt comprehension. In other words, the native speaker will know what is meant without undue effort, even if the sounds are not quite right and the foreign accent is thus fully audible. Computer speech, on the other hand, may be synthesized in such a way that there is no “foreign” accent, and yet the speech still does not sound quite right: few of us are ever deceived by computer speech into supposing that we are talking to a human being, and not a machine. Advances in technology enable such speech to be less “robot-like” than it was several years ago, but there is still far to go before a computer could carry on a conversation with

a person who would not even suspect that her interlocutor is actually a machine. And all this is to speak only of phonology, of the sound of the speech, without even entering the problem of diction, i.e. the choice of words.

The problem of fluency is not, of course, simply a matter of how quickly and accurately the L2 phonemes are realized. As the Shakespearean motto used at the head of this paper suggests, to “speak the speech trippingly on the tongue” also involves word choice; the actor can speak “trippingly” because the speech has been memorized, and Hamlet for his own reasons wants it spoken quickly. In any utterance there is a certain tolerance for a few brief pauses to select exactly the right word to be used, consciously considering several alternative words and rejecting all but one; depending on the circumstances, however, there can be only a few such pauses before the interlocutor loses patience and either terminates the conversation or tries to take control of it. This is true even of conversations between two native speakers, but in the case of the non-native speaker the problem is compounded. The non-native speaker must far more often make a careful and conscious choice between rival wordings, and far more often has difficulty in recalling the words. The frontal lobes are more heavily engaged in making lexical and grammatical decisions; as with any learned activity, the pattern of activation moves from anterior to posterior as the activity is learned or over-learned (Goldberg 2002). The frontal lobes, however, are relatively slow in real time, since their work usually involves making conscious decisions among alternative solutions. The more an innervation pattern becomes automatic, the fewer decisions are required, the longer the strings of neural commands that are activated only initially and then run their course, as in walking. If we had to think separately and consciously about the movement of every muscle in the body that is in motion when taking a single step, we would never be able to cross the street.

Of course, the *sine qua non* for fluent speech in L2 is eliminating or at least minimizing the interference of L1, i.e. the native language. To the extent that L1 must mediate between thought and speech in using L2 – i.e. the speaker thinks a thought in L1 and then translates it into L2 before uttering it – the computational demands on the brain are too high to enable the utterance to be delivered quickly and effortlessly. Experienced language teachers and users know this very well: in order to speak a language fluently, it is necessary to think in that language. This, indeed, is what underlies all the other conditions and criteria for fluency in SLA, and thus it is the only fully necessary and fully sufficient condition. As a native speaker of American English who has lived and worked in Poland for 15 years, I can testify to this personally: when I speak Polish, I think in Polish, which means that my “interior monologue” is in Polish. Switching languages involves shifting mental gears, i.e. changing not only the memory resources that are being tapped, but also the very nature of the thought that is being expressed. One need not accept the strong form of the Sapir-Whorf hypothesis to realize that this must be true. That is why simultaneous translation is so mentally taxing that inter-

preters must take long and frequent breaks, and even consecutive translation can be fully as exhausting as manual labor. Switching one's thoughts from language to language is accompanied by an energy cost in the brain; the more often and the more rapidly this is done, the higher the energy cost.

What must impress us as we consider these problems is the speed with which the brain must make the necessary adjustments in the speech apparatus to produce strings of phonemes that make comprehensible words in a given language, whether L1 or L2. We have been accustomed to thinking of computers as being far faster and more powerful than the human brain, a judgment which seems justified in respect to mathematical calculations, but a consideration of the problems raised here should give us pause. In a fluent sentence the realization of each phoneme is instantaneous, the duration measured in fractions of seconds, with more or less "white space" between them (including but not limited to word boundaries), and the sequencing is realized so quickly that there is no time for conscious decisions. The motor complexity of even the simplest linguistic utterances is impressive when compared to other body movements, which surely explains why so much of the neocortex undergoes activation when one speaks: not just the classic areas of Broca and Wernicke in the left hemisphere, which is where the most intensive neural activity takes place, but also extensive regions in both hemispheres. Rather than bemoan the slow and imperfect manner in which our brains solve arithmetic problems, we should pause for a moment to realize how much enormously complex work goes into my writing and your reading this one sentence. Most of this, of course, goes on below the threshold of consciousness, in the realm of what Jackendoff (2000) calls "fast processes." Having decided to utter a certain word, we do not then make separate decisions about each phoneme in that word, unless there is some uncertainty; analogously, many of the words that occur in a sentence do not appear as a result of a conscious decision in each case to say that word and not some other word.

From the neurolinguistic perspective, then, it would appear that acquiring fluency in a second language means shifting the balance of brain work in the target language from anterior to posterior. With increasing fluency the resources of conscious thinking are more and more freed from making syntactical, lexical and phonological judgments in real time, and can thus be used to "think ahead," focusing more on what is said than on how it said. The burden on the memory system shifts by the same token from explicit to implicit memory, as the decision to say something automatically (and subconsciously) fetches from memory the words needed to express that "something." There is still a lexical process of weeding out alternatives to pick just the right word, but in fluent speech more and more of this is done below the threshold of consciousness. Only the most salient and difficult concepts require a careful, conscious, deliberate choice of words. This is complicated in L2, however, by the fact that the words (and phonemes) are not so easily available to

these subconscious processes, and require far more conscious effort, which in turn slows down the process. The result is poor fluency: the speech in L2 does not flow, but takes on a stochastic character.

FLUENCY IN NEUROLINGUISTICS

In neurolinguistics, fluency is a quantitative measure of the rate of verbal output, the number of words produced in a given span of time. The “Boston School” of aphasiology, centered around the work of Norman Geschwind and Harold Goodglass, has popularized a classification of aphasia syndromes into two large categories, fluent and non-fluent, distinguished on the basis of the number of words the patient can utter in spontaneous speech (Goodglass & Kaplan 1983). Unlike the concept of fluency in the context of SLA, this kind of fluency pertains to the native language; moreover, it does not take into account the level of comprehension or comprehensibility in the speaker’s utterance. Thus a patient with “fluent” aphasia (to many, the term sounds like an egregious oxymoron) may say a great deal, as in the classic Wernicke’s aphasia, but what she says makes little sense (see some examples with a very useful discussion in Brown 1988). A “non-fluent” aphasia, on the other, may occur in a patient who understands spoken language without any particular difficulty, but cannot construct sentences more than a few words long, as in the classic form of Broca’s aphasia. The differentiation is made in the Boston Diagnostic Aphasia Examination, for example (Goodglass & Kaplan 1983), by showing the subject a picture, asking her to describe what she sees there, and counting the number of words used (regardless of their accuracy or sense).

A similar concept of fluency is often applied in the diagnosis of dementia (Pachalska 2007). Here the classic tests involve asking the subject to name as many animals as she can in one minute, or as many words beginning with a particular letter, or as many words with a specific number of letters, or some combination or variation of these tasks. Most patients in the early or early middle stages of Alzheimer’s dementia (DAT) do not perform below age-adjusted norms on these tests, especially if words that fail to meet exactly the stipulated criteria are also counted. Patients with semantic dementia or most forms of fronto-temporal dementia (FTD), on the other hand, almost always have difficulties producing more than a handful of words on any of these tests. In many FTD patients, a certain taciturnity noticed by family and friends at the beginning (often before any illness has been diagnosed) later turns into a syndrome called “logopenia,” when the patient’s verbal output is severely reduced, often confined to very short answers to simple questions, and finally ends in mutism, when there is no speech at all (Kertesz 2004).

What are we actually measuring when we use these kinds of tests? The task of naming as many things as possible that fall within a certain category is not a particularly “ecological” task, i.e. it is not something we are often asked to do, nor does it really resemble real-life tasks. On the other hand, it

does give us an objective, quantifiable measure of a difference that seems to have some clinical significance. Conversation with DAT patients differs obviously from conversation with FTD patients on this ground, that the latter speak in monosyllables or not at all, seldom initiating conversation or trying to maintain it if someone else takes the initiative, while many DAT patients are notoriously garrulous. The same applies to the difference that the clinician immediately perceives in patients with fluent or non-fluent aphasia. The results from the tests do in fact correlate strongly with these subjectively observed characteristics of speech, so they can be used to justify a diagnosis, or at least to specify a particular symptom that may or may not point to a diagnosis. This approach solves a practical, clinical problem, then, but it does not get us very far theoretically. We still do not necessarily know why this or that kind of brain dysfunction seems to produce more or less predictably this or that kind of disturbance in the fluency of speech.

IS THERE A “UNIFIED FIELD THEORY” OF FLUENCY?

As previously suggested, one possible solution to the problem posed at the beginning of this paper would be simply to assume that the word “fluency” means one thing in the context of SLA, and something else in neurolinguistics and neuropsychology. There is a risk, however, that in so doing we may be separating issues that can usefully be examined together, despite the obvious differences in meaning. Common sense may suggest, for example, that space and time are two separate things, but the realization that there is something called a “space-time continuum” has revolutionized physics in the 20th century. Common sense is not without its uses, but since it is almost invariably based on old theory (i.e. what seems commonsensical is usually whatever existing theory can explain without difficulty), it does not necessarily help much in developing new theory.

What do the speech of a Wernicke’s aphasic, an early DAT patient, and a competent speaker of a foreign language have in common, if each of them speaks “fluently”? At first glance, the best answer would seem to be: “almost nothing.” In Wernicke’s aphasia we expect to find neologisms, aphasic jargon, frequent paraphasias, resulting in speech that is barely if at all comprehensible. The speech of an early DAT patient is much less disordered by comparison, at least in terms of surface structure, but it is frequently filled with empty phrases, place savers, platitudes and the like, instead of substantive content. A fluent speaker of a foreign language, on the other hand, by definition does not produce disordered speech, but rather well-formed utterances in L2; fluency reaches 100% when the learner’s speech cannot be distinguished from that of a native speaker. In all three cases, however, there is a certain common ground, which is the tempo of speech:

- the speech of a Wernicke’s aphasic may be incomprehensible, but it flows along as though it were normal speech, which is precisely what justifies the term “fluent aphasia”;
- the speech of an early DAT patient may be empty or deficient in reasonableness, but again, it flows, often excessively (these patients not infrequently exhibit logorrhea instead of the logopenia of FTD);
- a well-formed sentence in L2, even if pronounced faultlessly, cannot be considered fluent if the tempo of speech is markedly slower than the norm for native speakers of that language.

The problem becomes even clearer if we look at it from the other side: what does the speech of a Broca’s aphasic or a patient with FTD have in common with the speech of an incompetent user of a foreign language? In each of these cases, nearly every word actually spoken results from an effort, not always successful, to find the right word and say it. In aphasics, as in persons who do not speak L2 very well, there are frequent agrammatisms and paraphasias resulting from inaccurate word choice, not to mention mispronunciations caused by imperfect control over the speech organs. The FTD patient differs from this picture in one important respect, that what little she does say usually seems to be accurate and fully sensible. Nevertheless, in this case, too, there is simply too little speech. Maintaining a conversation for any length of time requires a great deal of patience and forbearance from the interlocutor. When the patient begins a sentence and then falls silent for several seconds or even a minute, we do not know if she intends to complete the sentence or not.

In the case of the language learner, the reason for the lack of fluency, as suggested above, is usually the need to translate the utterance from the native language to the target language before it can be realized. Each word must be sought for and found before it can be said, which takes time, even when the learner is intelligent and remembers the grammatical and lexical rules she has been taught. Fluent speech, after all, *flows*, which even in the most literal sense means that the words come one after another without unnecessary interruption. Interruption, in turn, becomes necessary only when the target word does not follow all but automatically from what has already been said. This is the purpose of the grammatical, lexical, and morphological rules of a grammar, which are not so much markers of good and bad speech in the given language as they are constraints that allow the flow of language to be channeled in a particular direction. Without such constraints every word in an utterance could conceivably be followed by any other word existing in the lexicon of the given language, which is, of course, not the case.

Constraints, to be sure, are not determiners. Rather, they suggest what can follow, and what cannot follow without discontinuity. When one is competent in a given language (including the native language), at any given moment the number of possible utterances is limited only by the lexical and syntactical possibilities of that language and the user’s knowledge of it. When there is an intention to speak, from that *apeiron* of (nearly) infinitely many

possible utterances only a certain number are suitable. As I have elsewhere argued (MacQueen 2003), the way in which a sentence unfolds is more and more constrained by each successive word, which forces or prevents continuation of the sentence, permits or disallows various kinds of complements and modifiers, etc. Knowledge of the language allows both speaker and listener to be guided by these constraints in the desired direction, which is to comprehend the meaning. There must be heuristic decision nodes in this process, of course, but every successive word cannot be the result of a heuristic decision, or else the tempo of speech becomes so slow that before the sentence ends, no one remembers any longer how it began. This is the intuition that underlies phrase structure grammar: the speaker chooses a sentence, and then fills it with the requisite phrases, while the individual words needed to flesh out the phrase are generated automatically by the speaker's knowledge of the applicable lexical, syntactical, and morphological rules. The writer, of course, has the luxury of pondering each word, while the speaker must keep moving forward, relying on processes that once started do not need to be monitored. That is why fluency is a quality of speech, and can be predicated of writing only by extension, analogy, metaphor. By the same token, one does not choose phonemes heuristically: they are generated, or mandated, by the word that one wishes to say; this is what determines which phonemes begin the word, and then which phonemes must follow, in what order. The choice of the word is in large measure analogically constrained by the sentence that is being uttered, and that in turn is the realization of a certain proposition (in Hughlings Jackson's sense of the word, cf. Brown 1998) that both reflects and determines a particular mental state.

An analogy can be found in the way a melody or phrase works in a piece of music. Again, the composer begins with a virtually unlimited set of possible sounds, but the choice of an idiom, a genre, a key – all of these constrain what is possible. A melody is not simply an accidental concatenation of atomic sounds: it becomes a melody when it flows in a way that seems coherent, in a particular direction. Many, if not most of the successive notes seem determined, especially when the melody is being deliberately repeated, but even on its first occurrence the melody becomes cognizable when there seems to be a flow in a given direction. Again, as in a sentence, there are heuristic decision points, moments of possible discontinuity, when the composer (read: speaker) can do something other than what seems natural or even necessary. This effect cannot be repeated too often, however, or else the melody will lose its identity altogether: not only is it not *this* or *that* melody, it is no melody at all. Not every note is fully predictable on the basis of those that have already been played or sung, but many of them are, and perhaps most. Analogously, then, the speaker who overuses the privilege of deliberate discontinuity, i.e. allowing the listener to form and harbor expectations as to what is about to be said, and then saying something else instead, will not be understood. A certain number of discontinuities are possible, even desirable, even

necessary, but against the background of continuities, automatic sequences that are expected to run their course once they have been activated.

Another, perhaps even more interesting parallel can be found in quantum mechanics. A wave function in quantum physics represents a probability field, within which a particle may occur, though it need not occur at any given point; beyond the field determined by the wave function, on the other hand, the particle cannot occur. Quantum mechanics thus constitutes an inherently probabilistic, rather than deterministic system. At any moment before a moving particle actually appears in a certain place and allows itself to be observed and measured, there is a certain area, defined by the wave function, within which it may appear, and a rather larger area where (if the equation is correct) it will certainly not appear. Again, the wave function seems to be constraining the movement of the particle more than merely describing it, which accounts for one of the fundamental paradoxes of quantum mechanics: that an observed particle behaves differently from one that is not observed. Without going any further into this inordinately difficult area, we can surely notice at this point how language can be understood to behave according to similar laws and rules, which are probabilistic and not deterministic.

An example of this is the phoneme, which can reasonably be understood as a wave function. A very large number of speech sounds that are actually uttered, though they may exhibit sometimes very marked physical differences, will be perceived by native speakers of the same language as being “the same” phoneme or a permissible allophone. Some allophonic differences determine distinctions between class, regional, or generational accents, while others that exceed the boundaries of permissible sound variation are perceived as foreign and may cause a word to be incomprehensible to the native speaker. In other words, the phoneme distinguishes a larger or smaller (but in any event finite) number of allophones that will be heard as realizations of the phoneme. The analogy to a wave function is clear, where the sound that is actually uttered and heard at a given place and time represents the detected particle in a physical experiment. On the one hand, not every possible sound created by the speech organs can be a given phoneme, but on the other hand, it is not necessary for the physical prints of the sound to be identical to a certain fixed, i.e. deterministic pattern for two sounds made by two different speakers, or by the same speaker at two different moments in time, to be heard as being “the same.” The phoneme, then, is a probabilistic concept fully analogous to the wave function in quantum mechanics: *these* are the possible realizations, while *those* are not possible.

The same process of successively constraining the set of possible alternative solutions to a continuously smaller range until one and only one solution is finally realized takes place in lexical, morphological, and syntactical decisions (MacQueen 2003). Both the production and the comprehension of a sentence are probability functions, which may approach certainty, but would reach 100% only if a (hypothetical) ideal speaker said exactly what she wished to say, no more and no less with no ambiguities, and the (equally if

not more hypothetical) ideal listener heard and understood such a verbal message with no ambiguities, exactly as the speaker intended it. Even if such a thing could be posited as possible by an act of the imagination, of course, no putative occurrence could ever be checked or verified (how could we actually be sure that a message is “the same” in two different minds?), so it remains a purely speculative possibility of no practical importance. The rules of real language are probabilities, a series of bets made by speaker and listener simultaneously, where the number of wins relative to losses determines the level of comprehensibility and comprehension. This pertains to phonemes, lexemes, sentences, and texts in the same way, as the process is realized simultaneously in phonology, semantics, syntax, and pragmatics, though in different time scales depending on the temporal dimension of the respective events. This is not a sequential system (first phonology, then semantics, then syntax, then pragmatics, or any other hypothetical order), but a far more complicated arrangement in which there are mutual dependencies, simultaneities and discrepancies, not easily explained or charted.

From all this the point emerges clearly: time must be “chunked” in various ways (not in just one way!) for the mental processes involved in language (or, we may venture to suppose, any other mental act) to produce a coherent mental state (Brown & Pachalska 2003). This is a complex counterpoint that is possible only within a certain range of possible tempos. When things happen too quickly coordination is lost and everything becomes an indistinguishable blur; when the tempo is too slow, the gaps between the various elements and figures are too great, and even a simple straight line can begin to look stochastic. Current theory on the causes of stuttering have reached similar conclusions (Fava 2002), i.e. time within the language system is not organized in a simple linear fashion, but involves several clocks running simultaneously, in a highly complicated but not random system that is more probabilistic than deterministic, more heuristic and than algorithmic. When the phonological clock is out of synch with the lexical and syntactic clocks, one possible result is stuttering. The time of the sentence has moved forward too rapidly in relation to the time of phonological realization.

Again, the analogy with melody is useful: if the melody line is played too rapidly, there is only one sound, a chord and not a melody; if the tempo is too slow, it becomes a series of unrelated notes, the beginning of the phrase having been lost from memory long before the end is reached. Fluency in language, then, is quite the same thing. The phonemes, lexemes, and sentences must move slowly enough that they can be usefully differentiated, but rapidly enough that completion is possible before the beginning has been forgotten.

What determines, then, what is “too fast” and what is “too slow”? This is a brain problem, in a fairly obvious way. The central nervous system can perform within certain parameters that are biologically conditioned:

- “Too fast” means that stimuli follow upon each other so rapidly that they cannot be discriminated. The flight of a bullet cannot be traced visually unless it is slowed down on film; before the visual system (the eye, the

thalamus, the occipital lobes) can register the fact that the bullet is “there” it is already somewhere else. This is to say nothing of registering the fact in the form of a conscious perception. The brain sees nothing, neither the object nor its motion, because the successive “theres” cannot be actualized in the time made available by the capacities of the central nervous system.

- “Too slow” means that the capacity of working memory to keep objects in the field of consciousness over successive epochs of time (Brown 1996) is exceeded. We cannot see the motion of the minute and hour hands on a clock of less than monumental proportions because the successive, instantaneous “theres” are repeated with such little change between them that motion disappears. We must look away from the clock and then look back at it after some time has passed before we can deduce (rather than perceive) that the hand has moved.

What fluency does in language is to keep the objects of language (phonemes, lexemes, phrases, clauses, and so forth) moving at some rate that is neither too slow nor too fast. Failure in fluency, then, can be caused by two kinds of factors that adversely affect how the brain maintains this tempo:

- The brain itself may lose some or all of its capacity to keep objects moving in consciousness; in other words, the limits of “too fast” and “too slow” may be changed.
- The brain may not possess all the data it needs to form objects in the requisite time to keep them moving.

When the speaker’s knowledge of the language is imperfect, this means simply that the requisite objects (not only lexical, but syntactical and phonological as well) are not available, because they are not known at all, have been forgotten, or can be recovered only by conscious remembering, a mental process which takes far too long to keep the sentence moving. In dementia or aphasia, on the other hand, the brain itself is no longer able to create the requisite objects, or to string them into sequences, or both. In fluent aphasia, the former problem prevails; in non-fluent aphasia or logopenia (with many possible variants), the latter. Even so, however, the problems cannot easily be teased apart, which explains why the clinical picture in particular cases is almost never exactly what the various pathological schemes suggest. A solid object can be chopped into pieces that can be reassembled, or at least still bear some resemblance to the part of the whole they once were, but a fluid cannot.

CONCLUSION

The foregoing is the merest of sketches, and I am acutely, even painfully aware of the numerous leaps and gaps that could be filled and need to be filled. The aim here, however, was not so much to solve a particular problem, either in SLA or in neurolinguistics, but rather to step back from the day-to-day work in both areas and try to see a common, deeper problem. The acci-

dents of my own career have given me considerably direct experience in both foreign language learning (as teacher and learner) and in clinical work with brain-damaged patients, in various settings. What I have attempted to do here is to make these disparate bits of information into a sensible whole. There is always the danger in so doing of creating Dr. Frankstein's monster, but it may also be possible, sometimes, to discover things that would otherwise go unnoticed.

REFERENCES

- Brown, J.W. (1988). *The life of the mind. Selected papers.* Hillsdale, New Jersey, USA: Lawrence Erlbaum Associates.
- Brown, J. W. (1996). *Time, will and mental process.* New York: Plenum.
- Brown, J. W. (1998). *Psychoanalysis and process theory.* *Annals of the New York Academy of Sciences*, 843, 91-106
- Brown, J.W. & Pachalska, M. (2003). *The nature of the symptom and its significance for neuropsychology.* *Acta Neuropsychologica*, 1(1), 1-11.
- Chambers, F. (1998). *What do we mean by fluency?* *System*, 25(4), 535-544.
- Douglas, D. (2001). *Performance consistency in second language acquisition and language testing research: a conceptual gap.* *Second Language Research*, 17, 442-456.
- Fava, E., ed. (2002). *Clinical linguistics: theory and applications in speech pathology and therapy.* *Current Issues in Linguistic Theory* 227. Amsterdam and New York: John Benjamins.
- Goldberg, E. (2001). *The executive brain: frontal lobes and the civilized mind.* New York: Oxford University Press.
- Goodglass, H. & Kaplan, E. (1983). *The assessment of aphasia and related disorders.* 2nd ed. Philadelphia: Lea & Febiger.
- Jackendoff, R. (2000). *Fodorian modularity and representational modularity.* In: Y. Grodzinsky, L. P. Shapiro & D. Swinney (eds.), *Language and the brain. Representation and processing* (pp. 3–30). London: Academic Press.
- Kertesz, A. (2004). *Frontotemporal Dementia/Pick's Disease.* *Archives of Neurology*, 61(6), 969-971.
- MacQueen, B.D. (2003). *How a sentence unfolds.* *Acta Neuropsychologica*, 1(4), 382–406.
- Norris, J. M. & Ortega, L. (2000). *Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis.* *Language Learning*, 50, 417-528.
- Norris, J. M., & Ortega, L. (2003). *Defining and measuring acquisition.* In C. Doughty & M. H. Long (Eds.), *Handbook of second language acquisition.* Malden, MA: Blackwell.
- Pachalska, M. (2007). *Rehabilitacja neuropsychologiczna.* Lublin: UMCS Press.

Address for correspondence:

Prof. Bruce Duncan MacQueen
Institute of Psychology, University of Gdańsk,
Pomorska 68, 80-322 Gdańsk-Oliwa, Poland
e-mail: duncan.macqueen@medscimonit.com

Received: 2 September, 2007

Accepted: 28 December, 2007