Fluency, complexity and informativeness in native and non-native speech*

John Osborne
Université de Savoie

Individual speakers vary considerably in their rate of speech, their syntactic choices, and the organisation of information in their discourse. This study, based on a corpus of monologue productions from native and non-native speakers of English and French, examines the relations between temporal fluency, syntactic complexity and informational content. The purpose is to identify which features, or combinations of features, are common to more fluent speakers, and which are more idiosyncratic in nature. While the syntax of fluent speakers is not necessarily more complex than that of less fluent speakers, it is suggested that they are able to deliver content more efficiently through a combination of less hesitant speech and of lexical and syntactic choices that allow them to package information more economically.

1. Introduction

In what way is temporal fluency related to the informational content of speech? Is the ability to maintain a relatively high rate of speech, with a low percentage of hesitation time and long fluent runs, accompanied by a corresponding ability to package and deliver information efficiently? The answers to these questions are not necessarily straightforward, as can be illustrated by two short extracts from transcripts of different speakers recounting the same event (in references to corpus extracts, an initial N indicates a native speaker; all other extracts are from learners):

(1) s:o uh in this sequence we <can see> a child who is eating chocolate [//] a chocolate .
a:nd he is in a: zoo .
a:nd he: is showing (h)is chocolate to the: baby elephant .
a:nd the: [/] the animal wants to [*] the: chocolate . (PAROLE 002C)

(2) there we had an excellent commercial from the Netherlands it said on the bottom.
a:nd it's o:f a: little boy [//] (a) very //bratty boy it seems who has a Rolo .
and he: holds it up to <an elephant> [//] a young elephant .
a:nd the elephant looks like it really wants that Rolo but no the boy takes it for himself and eats it ! (PAROLE N14C)

By the end of their respective accounts, each of these speakers will have delivered the same quantity of information (as measured by the criteria which will be described below), but speaker 002 takes nine times longer to do so. Undoubtedly, speaker N14 is the more fluent of the two, but purely temporal fluency only partly accounts for this difference, since her speech rate, measured in words per minute, is just 3.3 times that of speaker 002. What else, then, accounts for the difference? Being able to produce more words per minute with fewer hesitations logically enables the speaker to deliver more information. But a higher speech rate is clearly not the sole cause of an ability to give information more rapidly, and there must be other factors helping speaker N14 to package information more efficiently. The purpose of this paper is to investigate what these factors might be, and how temporal fluency is related to the syntactic structure and informational content of oral production, both in first language (L1) and in second language (L2) production.

2. Corpus

The data are taken from the *PAROLE* corpus compiled at the Université de Savoie (Hilton et al. 2008). This corpus consists of 15–20 minute recordings of speakers of L2 English (L1 French and German), of L2 French (various L1s) and of L2 Italian (L1 French), together with recordings from native speakers of English, French and Italian carrying out the same tasks. The composition of the corpus is shown in Table 1.

Table 1. The *PAROLE* corpus

Language	n	L1	Gender
English	42	French (24), German (9), English NS (9)	F (35), M (7)
Italian	33	French (23), Italian NS (10)	F (25), M (8)
French	20	Spanish (5), Chinese (3), Swedish (2), Polish (1), English (1), French NS (8)	F (15), M(5)
Totals	95		F (75), M(20)

The recordings are transcribed in *CHAT* format (MacWhinney 2007) with detailed annotation for pauses (filled and unfilled), retracings (with or without

modification) and errors. The duration of all pauses was measured by hand, using the "Sonic mode" facility in *CLAN*.[1] This was preferred to the automatic detection of pauses available in tools such as *ELAN*, not only because it is more precise, but also because of the importance of filled pauses and the fact that many longer hesitation phenomena are in fact pausal groups consisting of one or more silent pauses combined with one or more filled pauses.[2] Back channelling (remarks not directly related to the central task, such as *Can I watch it again?*), use of L1 and metalinguistic comments such as *I don't know how to say it* are also annotated so that they can be excluded from certain calculations. For a fuller description of the corpus and the transcription guidelines, see Osborne (2007), Hilton (2008) and the *CHAT* manual (MacWhinney 2007).

All the speakers (95 in total) were university students, aged 18 to 25. The 68 L2 speakers also filled in a learner profile sheet and completed a number of additional tasks, including a vocabulary-size test, a non-word recognition test adapted, for English, from Gathercole & Baddeley (1996), and part of the *DIALANG* test.[3] They range in proficiency from A2 to C1 in the Common European Framework of Reference for Languages (Council of Europe 2001).

The analyses discussed here are based on the English and French components of the *PAROLE* corpus, since the Italian component is not yet fully transcribed and checked. The productions transcribed for the corpus come from five successive tasks, recorded in a single session for each speaker. Data from just two of the tasks will be considered here, since these were the tasks designed to elicit directly comparable content from speakers. Both are video-retelling tasks, summarised in Table 2, in which speakers were asked to watch a short (30–40 second) video film and then, without preparation, to recount what they had just seen.

Table 2. Video triggers

Task A duration 36s.	The sequence shows a group of men trying to hoist a large refrigerator into a top-floor window. Just as they are about to catch hold of the fridge, it slips out of its attachment and falls down onto the roof of a car parked at the bottom of the building.
Task C duration 35s.	The sequence shows a boy at a zoo, offering a chocolate to a baby elephant. Just as the elephant stretches out its trunk to take the chocolate, the boy snatches it back and eats it himself. The following scene shows the same person, now grown up, standing in a street watching a circus parade. He is about to eat a toffee when a large elephant in the parade taps him on the shoulder and then slaps him in the face with its trunk.

3. Measuring temporal fluency

The measurement of temporal features in speech production is itself a time-consuming activity. For earlier work on quantifying spoken fluency, see Goldman-Eisler (1968) and the collections of papers in Dechert & Raupach (1980) and Dechert et al. (1984). In these and in subsequent studies (Lennon 1990, Griffiths 1991, Towell et al. 1996, Cucchiarini et al. 2000, Towell 2002, Freed et al. 2004, Kormos & Dénes 2004, Mizera 2006), four main types of measures have been used: speech rate (in words or syllables per minute), quantity of pausing (as a percentage of total speech time or in average duration of pauses), mean length of runs (measured between two hesitation phenomena), and retracing. All of these measures were also used in measuring fluency in the *PAROLE* corpus. Given the degree of individual variation in oral production, it is unlikely that any single measure will provide a reliable indication of overall fluency. Some kinds of (dis)fluency behaviour, though, appear to be more idiosyncratic than others. In the data obtained from the *PAROLE* corpus, there is good agreement between measures of speech rate, percentage of hesitation time and mean length of runs. Each of these measures correlates well ($r > 0.800$, $p < 0.0001$) with each of the others, while the correlations are weaker ($r < 0.600$, $p < 0.002$) between retracing and the other measures. These first three measures — speech rate in words per minute, percentage of production time spent hesitating, and mean length of runs in words — were therefore combined into a composite 'fluency index', calculated in the following way:

For each of the three measures, the raw score for each subject was converted into a coefficient, by taking the average score of the native speaker group in each language as equal to 1 and the lowest learner score for the same language as 0.1. The coefficient for each individual speaker was calculated as 1 – (NS average – individual score) / ((NS average – lowest score) / 0.9). The coefficients for each measure (speech rate, percentage hesitation time and mean length of runs [MLR]) were then averaged to give an overall fluency index for each speaker. Table 3 shows examples for three selected speakers: a low-fluency learner (speaker 1), a high-fluency learner (speaker 2) and a native speaker (speaker 3).

Table 3. Sample fluency indices

	Speech rate		% hesitation time		Length of runs		Fluency index
Speaker	Score	Coeff	Score	Coeff	Score	Coeff	
1	42.01	0.11	74.60	0.10	2.20	0.14	0.12
2	152.48	0.80	23.70	0.99	8.20	1.03	0.94
3	213.30	1.17	21.20	1.03	7.60	0.94	1.05

This method was chosen in preference to other ways of standardising raw scores, such as calculating z-scores, principally because the results are more legible, in that scores close to 1.0 always indicate a performance comparable with that of an average native speaker. It also provides a score for each speaker which does not need to be recalculated if additional learners are included, although in the unlikely event that a new speaker proves to be less fluent than the least fluent speaker in the existing sample, his/her fluency index will fall below 0.1. For practical purposes of establishing an order of fluency among a set of subjects or constituting sub-groups, the fluency index and z-scores give almost identical results ($r = 0.999$, $p < 0.0001$).

The analyses presented here compare the productions of three sub-groups within the English and French components of the *PAROLE* corpus. The first group is composed of 12 low-fluency learners (10 learners of English and 2 of French), who are those with the lowest fluency index scores (ranging from .120 to .350). The second group are 12 high-fluency learners (8 of English and 4 of French), with fluency index scores ranging from .650 to .940. The third group is composed of 17 native speakers (9 of English and 8 of French), whose fluency index scores range from .690 to 1.350. As these scores indicate, the high-fluency learner group and the native speaker group overlap in terms of temporal fluency, some of the learners having a higher speech rate, with fewer hesitations, than certain native speakers. The measurements for the two target languages (English and French) are analysed together, because our principal concern was to investigate the common characteristics of fluency across languages, rather than to look at possible differences between languages. This is not to deny the role that such differences may have (see for example Carroll et al. 2000, Lambert & von Stutterheim 2003, Demol & Hadermann 2008), but learners' syntactic and informational choices may be influenced by their L1 or by their L2 to different degrees according to proficiency. To investigate these influences would require a larger and differently constructed corpus, with separate sub-groups at each proficiency level and for each combination of L1 and L2.

4. Measuring syntactic and informational content

In addition to the temporal measures described in the preceding section, more qualitative measures were also applied. These included accuracy, measured by the number of errors per 100 words, lexical range, measured by Vocd, an adaptation of type-token ratio which is less sensitive to sample size (see Malvern & Richards 1997), and syntactic and informational content. It is these last two measures that will be discussed now.

4.1 Syntactic content

The corpus was tagged using the *MOR* programme in *CLAN*, which provides part-of-speech and morphological information, but does not label higher-level components. This additional annotation was therefore done manually, a procedure which takes time, but not disproportionately so in relation to the other phases of transcribing and annotating an oral corpus. Syntactic components were labeled as either main clauses (independent, coordinate or superordinate), subordinate clauses (finite or non-finite), or verbless components (adverbial or prepositional phrases). Where necessary, incomplete syntactic units (for example clauses with a missing argument, such as *he is taunting*), were coded as clause fragments. The complete list of syntactic components coded is given in Table 4.

Table 4. Classification of syntactic units

Categories	Sub-categories
Main clauses	Independent clause 'Linked' clause (introduced by a connector) Coordinate clause Superordinate clause
Finite subordinate clauses	Relative clause Complement clause Adverbial clause
Non-finite subordinate clauses	*-ing* clause Past participial clause Infinitive clause
'Support' clauses	Existential clause (*there's…, it's…, I saw…*) Clause fragment (i.e. a matrix clause requiring a complement clause as one of the arguments, or any other clause lacking an argument)
Phrases	Adverbial phrase Prepositional phrase

Two types of clauses deserve particular mention, as they occur frequently in the data. The first are matrix clauses to which a complement clause is attached (*et là, on voit que l'éléphant petit est devenu grand* — "and then, you can see that the little elephant has become big"). The second are existential clauses of the type *il y a des éléphants qui passent sur la route* — "there are elephants that are walking along the road", which are frequently followed, as in this example, by a post-modifying clause or phrase. Quirk et al. (1985: 1408) describe existential clauses as "'presentative', in serving to bring something on to the discoursal stage deserving our attention". Among these are *There* + *BE* constructions, but also *I/we have…*, *One*

finds… and *There* + VERB constructions with verbs other than BE (which Ward et al. 2002: 1402–1403 call 'presentationals'). This function of "bring[ing] something on to the discoursal stage" is also fulfilled in our corpus by *It is…* constructions (e.g. *a:nd uh it's a young caïd (=bully) who have uh sweets in his hands*), which are probably specific to learner interlanguage, and, in the video-retelling tasks, by *I/we (can) see …* constructions (e.g. *we can see a child who is eating chocolate*). In the annotation, each of these types of clause has been given a specific coding, but since they serve a similar 'presentative' function, enabling the main information to be delivered in a dependent clause, they are also grouped, along with matrix clauses requiring a complement clause, into a category of 'support clauses'.

4.2 Informational content

For coding informational content, there is less established practice to draw on. Indeed it is doubtful whether it makes much sense to attempt an all-purpose definition of an information unit that could serve for any type of content. Procedures for analysing informativeness in discourse have been elaborated, notably for aphasia research, using CIU (correct information unit) analysis (Nicholas & Brookshire 1993) or the %IU (information units) metric (McNeil et al. 2001). There are also semantic coding systems such as SECS (Thorne 2004). All of these, however, have been elaborated with other research agendas in mind, and in terms of their usability or granularity of analysis are not well suited to our present purpose, which is to quantify how much and what kind of information speakers incorporate in their accounts, in relation to the information that they could have included. The procedure adopted is therefore to make a first run through all the transcripts in order to establish an inventory of information mentioned by at least one speaker. This is then used to make a check-list of the quantity and type of information that each speaker *could* have included in a maximally detailed description. Three types of 'information unit' are distinguished: (1) frames, which summarise an entire macro-event in a single statement; (2) micro-events, which break the macro-event down into discrete processes; and (3) adjuncts and attributes, which provide additional detail about the nature of the processes or the participants. Frames are thus introductory or concluding summaries of a macro-event, often beginning with an explicit cohesion marker, for example *and then years later the elephant decides to get his own back* or *so he gets his revenge*. Micro-events are typically formulated by unified predicates (Berman & Slobin 1994: 26), such as *the elephant approaches*, *he stretches out his trunk*, *he tries to take the chocolate*, *the boy takes back the chocolate*. Adjuncts and attributes are mostly adverbials (particularly prepositional phrases, e.g. *with his trunk*) or adjectives/adjectival phrases such as *an annoyong little (boy)*. The check-list is completed

for each transcript, noting the number of frames, micro-events and adjuncts/attributes actually mentioned by the speaker.

4.3 Utterance boundaries

For certain measures it is helpful to be able to calculate the number of words, of syntactic units or of information units per utterance, and the *CHAT* transcription format requires utterances to be separated by an utterance terminator. This supposes that one is able to identify, in spoken production, what constitutes a single complete utterance, something which in practice may be problematic.

In analyses of both written and spoken texts, the most frequently used unit of measure is the T-Unit, defined by Hunt (1965: 20) as "one main clause with all the subordinate clauses attached to it". Since it is not always easy to accommodate in this definition either the incomplete syntactic units or the run-on constructions that can occur in spoken production, Foster et al. (2000) propose an Analysis of Speech Unit (AS-Unit), which differs from a traditional T-Unit mainly in that it need not contain within it a complete clause. Detached sub-clausal units are a common feature of the spoken interactions analysed by Foster et al. (taken from an earlier study reported in Foster & Skehan 1996), but they are not frequent in the monologue productions analysed here, so the majority of utterances transcribed in the *PAROLE* corpus are in fact T-Units as defined by Hunt. An exception is made for certain coordinate clauses, when the temporal and prosodic characteristics suggest that they were produced as a continuous unit. Full details of the criteria for defining utterance boundaries, along with illustrative examples, are given in the *PAROLE* manual (Hilton 2008).

4.4 Combining the measures

Using the three types of measurement units just described (syntactic units, information units and utterances), it is possible to investigate the relationships of these units with each other and with speaking time, in order to quantify the syntactic and informational characteristics of speakers' productions. The measures based on these syntactic and informational units are summarised in Table 5.

The terms 'condensation' and 'granularity' are used here in a similar way to Noyau & Paprocka (2000), Noyau (2005) and Noyau et al. (2005). Condensation refers to the number of syntactic or information units in an utterance; granularity reflects the partitioning of situations into micro-events. For the analysis of the video summary tasks in *PAROLE*, syntactic condensation and information condensation have been measured separately, since the number of information units in an utterance does not necessarily map on to the number of syntactic units (although

Table 5. Syntax and content measures

	Syntax	Information
Rate	**syntactic rate** number of syntactic units per minute	**information rate** number of information units per minute
Density	**syntactic density** number of syntactic units per 100 words	**information density** number of information units per 100 words
Condensation	**syntactic condensation** number of syntactic units per utterance	**information condensation** number of information units per utterance
Ratio	**subordination rate** proportion of subordinate clauses to main clauses	**granularity** number of relevant statements made about each macro-event

the correlation between them is strong, $r = 0.894$, $p < 0.0001$). We have also measured *relative* granularity (the number of relevant statements made about each macro-event, as a proportion of the total number of potential micro-statements), in order to facilitate comparisons across tasks where the situations being described allow different degrees of partitioning into micro-events.

5. Results

5.1 Rate and density of syntax/information

The rate of syntactic and informational delivery is measured as the number of syntactic units or information units per minute. Tables 6 and 7 give the minimum, maximum and median rate of syntactic and information units, respectively, for each of the three groups: low-fluency learners (NNS low fluency), high-fluency learners (NNS high fluency) and native speakers (NS). Not surprisingly, both rates increase with fluency level ($r = 0.934$, $p < 0.0001$ and $r = 0.859$, $p < 0.0001$ respectively); one would expect that being able to produce more words per minute with fewer hesitations would indeed enable the speaker to produce more syntactic units and to deliver information more rapidly.

Examples (1) and (2) mentioned at the beginning of this paper suggest that more fluent speakers are able to deliver information not only more rapidly, but also more efficiently. Measures of syntactic and information density (i.e. the number of such units per 100 words — see Tables 8 and 9) show that density does

Table 6. Syntactic units per minute

	NNS low fluency	NNS high fluency	NS
Minimum	3.9	13.8	16.7
Maximum	12.1	31.8	38.2
Median	7.0	22.8	26.2

Table 7. Information units per minute

	NNS low fluency	NNS high fluency	NS
Minimum	3.2	7.3	11.7
Maximum	8.5	16.2	32.0
Median	4.9	14.4	20.8

Table 8. Syntactic units per 100 words

	NNS low fluency	NNS high fluency	NS
Minimum	12.9	10.7	11.6
Maximum	17.5	17.5	19.7
Median	14.4	15.2	16.5

Table 9. Information units per 100 words

	NNS low fluency	NNS high fluency	NS
Minimum	6.44	5.51	5.79
Maximum	13.11	11.86	17.42
Median	8.4	10.0	11.7

increase with fluency, but that the correlations are not strong: r = 0.339, $p = 0.0302$ for syntactic density; r = 0.330, $p = 0.0350$ for information density.

The three groups do, however, tend to produce different kinds of syntax. Table 10 shows the distribution of syntactic units, following the coding described in Section 4.1 above.

Table 10. Distribution of syntactic units

	NNS low fluency	NNS high fluency	NS
Phrases	57 (17.43%)	73 (15.24%)	135 (21.57%)
Support clauses	97 (29.66%)	123 (25.68%)	100 (15.97%)
Main clauses	102 (31.19%)	154 (32.15%)	243 (38.82%)
Subordinate clauses	71 (21.71%)	129 (26.93%)	148 (23.64%)

There are significant differences between the three sub-groups in the use of different syntactic structures ($\chi^2 = 35.18$, df = 6, $p < 0.001$). Analysis of individual cells indicates that the principal difference lies in the use of 'support clauses' by less fluent learners and by native speakers (residuals = 2.7 and 3.4 respectively). Other, less marked, differences between groups are that the native speakers produce proportionally more main clauses and phrases than both learner groups, while the fluent learners produce the most subordinate clauses and the fewest phrases.

The conclusion that can be drawn from this, at least tentatively, is that compared with less fluent learners, fluent learners and native speakers both produce speech that contains slightly more syntactic units, but that they tend to do so in different ways. Whereas the fluent learners rely more on subordination, incorporating dependent clauses into their utterances, the native speakers have a greater tendency to expand single-clause utterances with verbless phrases (for examples, see the discussion of lexical bundles in Section 6 below). More extensive analyses would be needed to determine how general this tendency is, and whether there is a level of fluency at which learners converge with native speakers in this respect.

5.2 Condensation

Condensation, as understood here, is a measure of the number of syntactic or information units per utterance, so that a high value for syntactic condensation reflects a tendency to produce syntactically complex utterances, and a high value for information condensation reflects an ability to package a lot of information into an utterance. Both correlate positively with fluency ($r = 0.777$, $p < 0.0001$ and $r = 0.733$, $p < 0.0001$ respectively); more fluent speakers produce utterances containing more syntactic units and more information (see Tables 11 and 12).

Table 11. Syntactic units per utterance

	NNS low fluency	NNS high fluency	NS
Minimum	1.04	1.68	2.38
Maximum	1.86	3.1	4.0
Median	1.3	2.1	2.8

Table 12. Information units per utterance

	NNS low fluency	NNS high fluency	NS
Minimum	0.53	0.86	1.31
Maximum	1.33	2.1	3.4
Median	0.8	1.4	1.9

This too is not surprising, but the exact nature of the relationship between syntactic structuring and information packaging is not so straightforward. Syntactic condensation logically allows the speaker to produce longer utterances and to accommodate more information in each utterance, as is reflected in the mean length of utterances for the three sub-groups (7.81 words for low-fluency learners, 13.88 words for high-fluency learners and 17.07 words for NS) and in a strong correlation between syntactic and informational condensation ($r = 0.896$, $p < 0.0001$). However, constructing longer, syntactically richer utterances does not by itself lead to more economical delivery of information. Measured by the amount of information delivered per 100 words, information density does not correlate strongly with syntactic condensation ($r = 0.406$, $p = 0.0085$), suggesting that an increase in hypotactic organisation does not always result in more efficient information packaging.

One reason for this may be that syntactically more elaborate utterances are sometimes used as circumlocutions to compensate for lack of lexical or other resources, as in example (3) below, where a low-fluency learner uses a difficult genitive relative (on the difficulty of genitive relatives, see Diessel & Tomasello 2005, Shirai & Ozaki 2007) to give the same information that a fluent learner (4) gives in a single noun phrase:

(3) and u:h the: [/] the man uh that [/] u:h that uh [/-] who u:h is the car↑ is u:h very u:h [/-] is not happy. (PAROLE 004A)

(4) and I think the car's owner was screaming he was uh raising his hands and he was uh screaming +"/. (PAROLE 025A)

The same phenomenon appears in verbal constructions, as illustrated by the gradually elaborated causative construction used in example (5), to give essentially the same information as that encoded in (6) by the single verb "hoist", which is used by half of the native speakers in the corpus but by none of the learners, perhaps not surprisingly since "hoist" is a relatively infrequent word in English, lying outside Kilgarriff's (1996) frequency list of the 6,000-odd words occurring 800 times or more in the *British National Corpus*.

(5) so <u:m #> [#1_068] there [*] <u:m # &=bouche> [#1_312] two or three people↑ [*] #0_888 at a wi:ndo:w [*] on the: second [*] floor o:f [///] #0_424 well I don't know which floor .
#0_546 a:nd they are trying to: [/] <um # &=rire> [#4_086] +//.
+, &=rire to <um: # &=bouche #> [#3_517] <0subj don't know> ["] .
#0_435 okay: they try to: <u:m # &=snap # &=bouche> [#4_969] <have a refrigerator reaching> [*] the window [//] +/.
+, to <have the refrigerator going> [*] #0_285 through the window↑ (PAROLE 020A)

(6) so there's a crane hoisting a refrigerator #0_279 up into: uh it looked like a [/] u:m [#0_418] a: third storey window . (PAROLE N11A)

5.3 Ratio

Granularity is another dimension to information content. Largely irrespective of how fluently or efficiently information is delivered, a speaker may choose to present a situation as a single macro-event or to break it down into a series of micro-events.

Examples (7) to (9) below, all from native speakers, illustrate how granularity can vary between speakers of (presumably) comparable language proficiency.

Low granularity: English

(7) (o)kay I saw a: [//] something white being hoisted up to the top of a: building with some men at the top waiting at the window with open arms to receive it .
+^ 0subj turned out to be a: fridge I think .
a:nd just as they got their hands on it it slipped out of the: [/] the hold↑ [//] the: [/] the rope that was around it and landed on a: green car beneath it &=rire ! (PAROLE N03A)

Low granularity: French

(8) je [/] je pense que c'est un emménagement de: frigo (en)fin ils veulent mettre un frigo dans une maison.
donc ils le montent à l'aide d'une grue.
mais uh arrivé à la fenêtre donc pour le faire passer dans la maison be:n ils ont pas réussi à le retenir et il est tombé <dans une voiture> [//] sur une voiture.
"I think they're moving in a fridge they want to take a fridge into a house so they lift it up with the help of a crane
but uh when it gets to the window to bring it into the house they didn't manage to hold it back and it fell <into a car> onto a car." (PAROLE N48A)

High granularity: English

(9) we:ll u:m there were these guys trying to hoist a big fridge uh up into a window.
about three guys in the window a:nd there's one guy on the ground.
and they almost got it in it was very [?] up to the window and they were reaching for it and the:n it fell .
[oh no !]
and of course there was a car right under the window +/.
+, 0subj] crushed the car +.

(a)nd there's the guy who's like &ge gesturing madly with his ha:nds and the fridge is just ruined and the car is also ruined .
[oh no !]
it was actually pretty funny &=rire !
+<but not for them I guess . (PAROLE N12A)

When one compares speakers from all three groups, there is a moderate correlation between granularity and fluency ($r = 0.625$, $p < 0.0001$) and no correlation ($r = 0.0451$, $p = 0.7797$) between granularity and information density, measured by the number of information units per 100 words. Table 13 shows the granularity of narratives (i.e. the number of relevant statements made, as a percentage of the total number of potential micro-statements) by the three groups of speakers.

Table 13. Granularity

	NNS low fluency	NNS high fluency	NS
Minimum	16	30	24
Maximum	40	48	56
Median	26	37	40

It would seem, then, that granularity is a question of individual choice rather than a characteristic of more or less fluent speech. To a large extent this is the case; there is no significant difference in granularity between the native speaker sub-group and the more fluent learners. In the whole corpus, the speaker with the highest granularity and the speaker with the lowest granularity, both learners, have almost exactly the same overall fluency index (0.53 and 0.55 respectively). Unlike most other measures, granularity measures for each of the two video-description tasks do not correlate strongly ($r = 0.481$); the same speaker may opt for a high degree of granularity in one task but not in the other. However, there is a significant difference in granularity between the two learner groups, with the low-fluency group having lower granularity than the high-fluency group, so while granularity does not appear to be a general characteristic of fluent speech, it is possible that at lower levels of proficiency the degree to which speakers can exercise their choice of granularity is limited by lack of linguistic resources.

Similarly, there is considerable individual variation in subordination rate (the percentage of subordinate clauses in relation to the total number of clauses produced by the speaker — see Table 14), and no clear correlation between subordination rate and fluency ($r = 0.145$, $p = 0.3656$). Four of the least fluent learners do, however, have an exceptionally low subordination rate (less than half the median rate for native speakers) suggesting that this too may be a result of linguistic limitations rather than an individual choice.

Table 14. Subordination rate

	NNS low fluency	NNS high fluency	NS
Minimum	5.88	17.95	18.52
Maximum	43.48	39.13	47.37
Median	21.2	29.1	30.8

6. Hesitation patterns

For most speakers, subordination rate and granularity appear to be largely a question of individual choice, not closely related to overall fluency. However, there are some interesting differences between groups in the degree to which they pause before particular types of information unit. This can be seen by looking at the three longest hesitations produced at a phrase, clause or utterance boundary by speakers in each of the three sub-groups and listing the types of information unit that follow. All three types of information unit — frames, micro-events and adjuncts/attributes — appear after long pauses in each of the sub-groups, but in differing proportions. The most interesting differences concern framing units, those which summarise a macro-event or which serve as an introduction or conclusion to the macro-event, and are therefore indicators of the way items of information are packaged into larger units. In comparison with the more fluent learners and with native speakers, the low-fluency group produce fewer of these framing units and have more major hesitations — 2 seconds or more — before them. Overall, 70% of the framing units produced by the low-fluency group are preceded by a long hesitation (between 2 and 10 seconds), as in example (10), compared with 21% for the high-fluency learners and 13% for the native speakers.

(10) and <u:m # u:m #> [#5_321] the elephant <u:h # &=bouche # uh #> [#9_766] we(ll) takes [*] his u:h vengance @n [*] (PAROLE 15C)

At the next level down, there are no clear differences in the association of major hesitations with specific micro-events. Those events that appear more frequently after a long pause are also those which are more often mentioned. In fact, the main difference between groups lies not in whether the formulation of a particular micro-event is preceded by a major hesitation, but in whether the event is mentioned at all. In the first summary task, for example, 77% of the native speakers mention the micro-event "a fridge is lifted/raised/hoisted by a crane" / "un frigo est tracté/soulevé par une grue". The same micro-event is mentioned by 50% of the high-fluency learners, but by only 17% of the low-fluency learners. This is probably a case of topic avoidance motivated by lexical limitations, in all the cases where less

fluent learners do mention this event, it is accompanied by a metalinguistic comment:

(11) [#2_656] j' ai vu devant &u <u:h #> [#0_685] une [*] immeuble <# u:m # &=rire # u:h> [#7_286] une [/] une machi:ne [*] uh <je sais pas comment dire> ["] .
"I saw in front of uh a building um a machine uh <I don't know how to say it>" (PAROLE 415A)

In the same task, the micro-event "the fridge slips out of the rope" / "il tombe du câble" is mentioned by 35% of the native speakers, usually accompanied by some hedging or lexical groping, as in (12) and (13):

(12) et il est lâché <de sa:> [/] #0_557 de sa: #0_674 nacelle↑ <je sais pas> ["] "and it is dropped from its [/] from its cradle <I don't know>" (PAROLE N40A)

(13) a:nd just as they got their hands on it it slipped out of the: [/] #0_343 the hold↑ [//] the: [/] the rope that was around it. (PAROLE N03A)

Given these lexical uncertainties from native speakers, it is perhaps not surprising that this micro-event is not mentioned by any of the learners, who conflate it into a more coarse-grained description, "and then it fell down". Similarly, in the "elephant" video-retelling task, the micro-event "he is watching a parade/procession" / "il regarde un défilé" is mentioned by 82% of the native speakers and by 83% of the more fluent learners (sometimes with circumlocutions or lexical approximations, e.g. *watching a carnival or something like that*), but is mentioned by only 25% of the less fluent group, often with more remote lexical approximations (*un homme qui: [/] […] qui regarde uh une fête à côté de la route*; *he's u:m I think er in a zoo or in a: party with u:h animals*). It is probably this kind of avoidance which is responsible, at least partly, for the differences in granularity between the two learner groups; certain micro-events are probably omitted by less fluent speakers not for reasons of information structuring, but simply because they are unable to encode them.

A similar lack of resources may also limit choices of syntactic condensation. As we have seen, the low-fluency sub-group make greater use of what we call 'support' clauses. These are main clauses that would not normally constitute a complete syntactic structure by themselves, but are used to attach a dependent clause, usually a complement clause or a restrictive relative, which actually carries the main informational weight of the utterance. One of their main functions is to move the narration on step by step, by means of existential clauses, such as "there is an x", followed by "who/which…". These syntactic structures enable the speaker to start from the familiar — reintroducing given information and previously used language — and to

add new information in a subordinate clause. Examples (14) and (15) illustrate this, compared with (16), in which the same information is furnished in a single clause.

(14) <# uh> [#0_511] aprè:s <e:m &=bouche #> [#1_161] il y a un homme que [*] c' est [*] le [*] enfant↑ [*] mais <quand il &è:t> [/] quand il étai:t &grã <# eu:h> [#0_743] &j uh <plus vieux> [//] oui [“] quand il était [*] un adulte [*] .
“uh afterwards there is a man that he is the child but <when he wa> when he was big uh <older> yes when he was an adult” (PAROLE 409C)

(15) #0_291 and <u:h # &=bouche> [#1_793] after [*] uh we see: uh [#0_360] the: [*] [/] #0_731 the young kid u:h w(h)o [*] [/] <u:h #> [#1_202] w(h)o have [*] u:h [#0_644] grown up . (PAROLE 001C)

(16) <&=bouche #> [#0_830] so then it shows later on in the boy's life. (PAROLE N15C)

Existential clauses are structurally simple, and in the native speaker group are those which have the shortest mean clause-initial pauses. In the learner groups, however, the mean length of clause-initial pauses is greater for existential clauses than for main clauses or subordinate clauses, possibly because learners particularly favour existential constructions in cases where they are uncertain how to encode the upcoming information. As we saw in Table 10 above, learners produced proportionally fewer main clauses and phrases than the native speakers, and syntactically elaborate utterances may be used by learners as circumlocutions to compensate for lack of linguistic resources, as in examples (3) and (5) cited above. Circumlocutions of this kind are one reason why syntactic condensation does not necessarily result in more efficient packaging of information. Whereas more fluent speakers can make choices from a variety of lexical and syntactic means, less fluent learners often appear to proceed in a step-by-step fashion, by successive increments.

In addition to the duration and position of pauses, it is instructive to look at what lies in between. The distribution of pauses can be used, for example, as a way of validating multiword expressions, to complement automatic extraction of n-grams (see Dahlmann & Adolphs 2007). It is also interesting to start from the n-grams themselves, and ask in what way they may contribute to the fluency and to the content of oral production. Table 15 shows, for each group of speakers, the number of different three-word bundles occurring at least three times in their productions (followed in brackets by the number of tokens). The bundles are classified as 'filled' bundles (those in which one of the components is a filler such as *uh* or *um*), presentative bundles (e.g. *we can see*, *there is a*), stance bundles (e.g. *I don't know*, *I think it's*), textual bundles (e.g. *and then he*, *and then it*), or referential bundles (e.g. *with his trunk*, *through the window*).

Table 15. Types (and tokens) of three-word bundles

	NNS low fluency	NNS high fluency	NS
'Filled' bundles	22 (76)	2 (7)	0 (0)
Presentative bundles	5 (33)	8 (33)	0 (0)
Stance bundles	3 (18)	5 (20)	1 (3)
Textual bundles	0 (0)	0 (0)	2 (7)
Referential bundles	5 (16)	10 (36)	22 (77)
Total	**35 (143)**	**25 (96)**	**25 (87)**

In the low-fluency learner group, filled bundles predominate; indeed the third most frequent cluster used by this group is composed exclusively of fillers, the hesitation bundle *um, bouche, er* ("bouche" is the coding used in *PAROLE* to transcribe non-phonemic filler sounds such as tongue clicks and lip smacks). This group therefore produce relatively few 'genuine' lexical bundles, except insofar as fillers such as *er* and *um* can be considered to have lexical or semi-lexical function (see Clark & Fox Tree 2002). The lexical bundles that they do produce are most frequently presentative, as might be expected from the discussion above. The high-fluency learner group produce a similar number of presentative bundles, but far fewer filled bundles, and a greater number of referential bundles. There are no recurrent presentative bundles in the native speaker group (with the exception of *there was a*, which occurs twice), confirming that these formulae are a characteristic of less proficient speakers. The vast majority of native speaker bundles are referential, the most frequent being prepositional phrases, such as *with his trunk*, *in his mouth*, *through the window*, *on the shoulder*. Among the learners, the more fluent group produce just three different prepositional bundles (*in a zoo*, *on a car* and *into the house*), and the less fluent group only one, namely *in the street*. It is probable that the ability to produce referential bundles of this kind is one of the factors that enable more fluent speakers not only to attain a higher rate of speech, but also to deliver information more efficiently.

7. Conclusion

The "smooth flow of language" referred to in the C1-level descriptors of the Common European Framework of Reference (*CEFR*) for Languages (Council of Europe 2001: 28) is an important characteristic of fluent speech, but is only the first of four kinds of fluency described by Fillmore (2000: 51), along with the ability to talk in coherent, reasoned and "semantically dense" sentences, to have appropriate things to say in a wide range of contexts and to be creative and imaginative in

language use. Lennon (2000: 26) uses no fewer than five adjectives to formulate a working definition of fluency as "the rapid, smooth, accurate, lucid, and efficient translation of thought or communicative intention into language". Here, we have concentrated on the ways in which the first two of these qualities (rapid, smooth speech) may be related to the last (efficient speech, as measured by the amount of information delivered), and to the kinds of syntactic structuring used to combine bits of information into an utterance. Accuracy too can be an important factor in the perception of fluency (see for example Kormos & Dénes 2004, Brand & Götz this volume), and the error rate of our speakers correlates well with fluency ($r = -0.823$, $p < 0.0001$). The relations between accuracy and fluency have not been investigated further here, principally because the quantification of errors raises numerous questions. For example, should pronunciation errors be included? If so, what degree of divergence constitutes an error? Should recurrent errors be counted repeatedly or just once? Accuracy, like fluency, is multi-faceted, and a more detailed investigation, requiring a combination of measures, is outside the aims of the present study, as is a more systematic discussion of the relationship between 'fluency' and 'proficiency', although comparison of learners' fluency measures with their results on tests of structure, vocabulary and comprehension do suggest that more fluent learners are frequently more proficient in other respects also.

Most of the qualities mentioned above are incorporated, in one way or another, into the descriptors used in evaluating oral proficiency in foreign languages. The *CEFR* C1-level descriptors mentioned above also refer to "clear, detailed descriptions" and to "well-structured speech, showing controlled use of organisational patterns, connectors and cohesive devices". Similarly, the *ACTFL* guidelines (American Council on the Teaching of Foreign Languages 1999) describe "Advanced-High" speakers as those who "are able to consistently explain in detail and narrate fully and accurately in all time frames" and "use precise vocabulary and intonation to express meaning and often show great fluency and ease of speech". These descriptors aim to capture experienced assessors' perceptions about the nature of oral performance at different proficiency levels. However, they remain essentially subjective, and are intentionally independent of any specific language. It is thus useful to be able to check them against analyses of the features that actually appear in learners' productions, in order to be more specific about what constitutes "ease of speech" or "smooth flow of language", about the extent to which this varies (between individuals, between tasks, between languages) and about its relationship with other aspects of oral production, such as accuracy, range, and the structure and content of speech. It has not been possible to cover all of these aspects in this short study, but among the tentative lessons that can be drawn from the preceding discussion are the following:

Criteria such as "can give clear, detailed descriptions" (*CEFR*) or "are able to consistently explain in detail and narrate fully and accurately" (*ACTFL*) may need to be handled with caution, given the extent to which granularity can vary between individuals and between the same speaker's performance on two even quite similar tasks.

The greater use of what we have called framing units by more fluent learners and by native speakers suggests that, rather than the ability to provide detail, it is often the capacity to introduce, synthesise and conclude a description that characterises more fluent speakers. This is in accordance with one of the characteristics specifically mentioned in the *CEFR* C1-level descriptors: "rounding off with an appropriate conclusion".

The use of circumlocutions ("allowing gaps to be readily overcome with circumlocutions" [*CEFR* C1], "confident use of communicative strategies, such as paraphrasing, circumlocution, and illustration" [*ACTFL* Advanced-High]) is double-edged. They do indeed allow gaps to be overcome — although hesitation patterns suggest that this is not always achieved readily — but circumlocutions render the delivery of content less efficient, and are more characteristic of less fluent speakers. Individual variation also comes into play in deciding whether to attempt a circumlocution or to avoid a topic altogether.

References to "well-structured speech" (*CEFR*) or to "structured argument" (*ACTFL*) are in practice quite difficult to interpret objectively, since structuring can be observed at various levels, and the relations between syntactic structuring, textual structuring and information structuring are complex, as are their respective relationships with language proficiency.

The preceding remarks apply to descriptors for higher levels of proficiency, C1 and Advanced-High respectively, but the analysis of fluency, syntax and content in oral productions has similar implications at other levels too. It is hoped that such analyses will contribute to a better understanding of what typifies learner productions at each level, and will lead to possible applications in the benchmarking of representative samples.

Notes

* This research was carried out as part of the *PAROLE* project at the Université de Savoie, and would not have been possible without the work of the other members of the *PAROLE* team. Particular thanks are due to Heather Hilton, whose work on pause distribution was invaluable for many of the analyses presented here. I should also like to thank the editors of this volume and two anonymous reviewers for their helpful comments.

1. The *CLAN* programme ("Computerised Language Analysis") was developed as part of the CHILDES project. It is available from the CHILDES website: http://childes.psy.cmu.edu/.

2. *ELAN* is a video and audio annotation tool developed at the Max Planck Institute for Psycholinguistics, and can be downloaded through their Language Archiving Technology Portal: http://www.lat-mpi.eu/ (accessed January 2010).

3. At the time the tests were made, *DIALANG* was run from the *DIALANG* website server (http://www.dialang.org/), which is presently unavailable. *DIALANG* can currently be used from the test server at Lancaster University: http://www.lancs.ac.uk/researchenterprise/dialang/about (accessed January 2010).

References

American Council on the Teaching of Foreign Languages. 1999: online. *ACTFL Proficiency Guidelines — Speaking*. Available at: http://www.actfl.org/i4a/pages/index.cfm?pageid=3325 (accessed August 2010).

Berman, R. & Slobin, D. 1994. *Different Ways of Relating Events in Narrative: A Cross-linguistic Developmental Study.* Hillsdale NJ: Lawrence Erlbaum.

Carroll, M., Murcia-Serra, J., Watorek, M. & Bendiscioli, A. 2000. "The relevance of information organization to second language acquisition: The descriptive discourse of advanced adult learners of German". *Studies in Second Language Acquisition*, 22, 441–466.

Clark, H. & Fox Tree, J. 2002. "Using *uh* and *um* in spontaneous speaking". *Cognition* 84 (1), 73–111.

Council of Europe. 2001. *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Cambridge: Cambridge University Press.

Cucchiarini, C., Strik, H. & Boves, L. 2000. "Quantitative assessment of second language learners' fluency by means of automatic speech recognition technology". *Journal of the Acoustic Society of America*, 107 (2), 989–999.

Dahlmann, I. & Adolphs, S. 2007. "Pauses as an indicator of psycholinguistically valid multiword expressions (MWEs)?". *Proceedings of the Workshop on A Broader Perspective on Multiword Expressions*. Prague: Association for Computational Linguistics, 49–56.

Dechert, H. W. & Raupach, M. (Eds.) 1980. *Temporal Variables in Speech*. The Hague: Mouton.

Dechert, H. W., Möhle, D. & Raupach, M. (Eds.) 1984. *Second Language Productions*. Tübingen: Gunter Narr.

Demol, A. & Hadermann, P. 2008. "An exploratory study of discourse organisation in French L1, Dutch L1, French L2 and Dutch L2 written narratives". In G. Gilquin, S. Papp & M. B. Díez-Bedmar (Eds.), *Linking Up Contrastive and Learner Corpus Research*. Amsterdam: Rodopi, 255–282.

Diessel, H. & Tomasello, M. 2005. "A new look at the acquisition of relative clauses". *Language*, 81 (4), 882–906.

Fillmore, C. 2000. "On fluency". In H. Riggenbach (Ed.), *Perspectives on Fluency*. Ann Arbor: University of Michigan Press, 43–60. First published in C. Fillmore, D. Kempler & W. Wang

(Eds.), 1979, *Individual Differences in Language Ability and Language Behavior*. New York: Academic Press, 85–101.

Foster, P. & Skehan, P. 1996. "The influence of planning on performance in task based learning". *Studies in Second Language Acquisition*, 18 (3), 299–324.

Foster, P., Tonkyn, A. & Wigglesworth, G. 2000. "Measuring spoken language: A unit for all reasons". *Applied Linguistics*, 21 (3), 354–375.

Freed, B., Segalowitz, N. & Dewey, D. 2004. "Context of learning and second language fluency in French: Comparing regular classroom, study abroad, and intensive domestic immersion programs". *Studies in Second Language Acquisition*, 26 (2), 275–301.

Gathercole, S. & Baddeley, A. 1996. *The Children's Test of Nonword Repetition*. London: Psychological Corporation.

Goldman-Eisler, F. 1968. *Psycholinguistics: Experiments in Spontaneous Speech*. New York: Academic Press.

Griffiths, R. 1991. "Pausological research in an L2 context: A rationale, and review of selected studies". *Applied Linguistics*, 12 (4), 345–362.

Hilton, H. 2008: online. "Le corpus PAROLE: Architecture du corpus et conventions de transcription". Available at: http://talkbank.org/BilingBank/ (accessed January 2010).

Hilton, H., Osborne, J., Derive, M.-J., Suco, N., O'Donnell, J. & Rutigliano, S. 2008: online. *Corpus PAROLE*. Available at: http://talkbank.org/BilingBank/ (accessed January 2010).

Hunt, K. 1965. *Grammatical Structures Written at Three Grade Levels. NCTE Research Report No. 3*. Champaign Ill.: NCTE. Available from ERIC: http://eric.ed.gov (No. ED113735, accessed January 2010).

Kilgarriff, A. 1996: online. "BNC database and word frequency lists". Available at: http://www.kilgarriff.co.uk/bnc-readme.html (accessed January 2010).

Kormos, J. & Dénes, M. 2004. "Exploring measures and perceptions of fluency in the speech of second language learners". *System*, 32, 145–164.

Lambert, M. & von Stutterheim, C. 2003. "La subordination dans les récits d'apprenants avancés francophones et germanophones de l'anglais". *AILE*, 19, 41–69.

Lennon, P. 1990. "Investigating fluency in EFL: A quantitative approach". *Language Learning*, 40 (3), 387–417.

Lennon, P. 2000. "The lexical element in spoken second language fluency". In H. Riggenbach (Ed.), *Perspectives on Fluency*. Ann Arbor: University of Michigan Press, 25–42.

MacWhinney, B. 2007: online. *The CHILDES Project: tools for analyzing talk — electronic edition*. Available at: http://childes.psy.cmu.edu/manuals/chat.pdf and /clan.pdf (accessed January 2010).

Malvern, D. & Richards, B. 1997. "A new measure of lexical diversity". In A. Ryan & A. Wray (Eds.), *Evolving Models of Language*. Clevedon: Multilingual Matters, 58–71.

McNeil, M., Doyle, P., Fossett, T., Park, G. & Goda, A. 2001. "Reliability and concurrent validity in the information unit scoring metric for the story retelling procedure". *Aphasiology*, 15 (10/11), 991–1006.

Mizera, G. J. 2006: online. *Working Memory and L2 Oral Fluency*. Unpublished PhD dissertation, University of Pittsburgh. Available at: http://etd.library.pitt.edu/ETD/available/etd-04262006-124945/unrestricted/Mizera.Dissertation.pdf (accessed January 2010).

Nicholas, L. & Brookshire, R. 1993. "Quantifying connected speech of adults with aphasia". *Journal of Speech and Hearing Research*, 36, 338–350.

Noyau, C. 2005. "Le lexique verbal dans des restitutions orales de récits en français L2: Verbes de base, flexibilité sémantique, granularité". In F. Grossmann, M.-A. Paveau & G. Petit (Eds.), *Didactique du lexique: Langue, cognition, discours*. Grenoble: Ellug, 65–84.

Noyau, C. & Paprocka, U. 2000. "La représentation de structures événementielles par les apprenants: Granularité et condensation". *Roczniki Humanistyczne*, 48 (5), 87–121.

Noyau, C., de Lorenzo, C., Kihlstedt, M., Paprocka, U., Sanz Espinar, G. & Schneider, R. 2005. "Two dimensions of the representation of complex event structures: Granularity and condensation. Towards a typology of textual production in L1 and L2". In H. Hendriks (Ed.), *The Structure of Learner Varieties*. Berlin: Mouton de Gruyter, 157–201.

Osborne, J. 2007. "Investigating L2 fluency through oral learner corpora". In M. C. Campoy & M. J. Luzón (Eds.), *Spoken Corpora in Applied Linguistics*. Frankfurt am Main: Peter Lang, 181–197.

Quirk, R., Greenbaum, S., Leech, G. & Svartvik, J. 1985. *A Comprehensive Grammar of the English Language*. London: Longman.

Shirai, Y. & Ozaki, H. 2007. "Introduction" (Special issue: *The acquisition of relative clauses and the Noun Phrase Accessibility Hierarchy: A universal of SLA?*). *Studies in Second Language Acquisition*, 29 (2), 155–167.

Thorne, J. 2004: online. *The Semantic Elaboration Coding System (SECS): Training Manual.* Available at: http://students.washington.edu/jct6/Handbook.htm (accessed January 2010).

Towell, R. 2002. "Relative degrees of fluency: A comparative case study of advanced learners of French". *IRAL*, 40 (2), 117–50.

Towell, R., Hawkins, R. & Bazergui, N. 1996. "The development of fluency in advanced learners of French". *Applied Linguistics*, 17 (1), 84–119.

Ward, G., Birner, B. & Huddleston, R. 2002. "Information packaging". In R. Huddleston & G. Pullum (Eds.), *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press, 1365–1447.