The role of formulaic sequences at different stages of foreign language learning

Florence Myles

Vitoria, 16th May 2014
Outline

• Formulaic sequences: what are they?
  – In native speakers
  – In second language learners

• Research agenda

• Identification issues
  – In beginner learners
  – In advanced learners

• The studies

• Use of FS
  – In beginners
  – In advanced stages

• Implications for SLA

• Pedagogical implications
Formulaic sequences: what are they?

• Speaker external definition
  – Sequence of morphemes commonly found together in native speaker corpora
    • E.g. *will you marry me?* rather than *would you like to get married with me? Nice day today* rather than *the weather is good today*
  – Sequence of morphemes which is a single semantic unit which cannot be generated from its individual constituents
    • E.g. *kick the bucket; by and large; thanks for having me*

• Speaker internal definition
  – Same definition but for a given individual
  – Processing shortcut for that particular individual
  – What is formulaic in one individual is not necessarily formulaic in corpora or in the target language, and vice versa, e.g.
  – In the studies reported here psycholinguistic definition (Wray 2002; Myles et al. 1998)
Examples of FS

• In native speakers
  – Idiomatic expressions e.g. *it’s raining cats and dogs; to kick the bucket; hit and run*...
  – Fillers e.g. *I think; you know; having said that*...
  – Preferred expressions (within an individual or community) e.g. *tha knows* in Sheffield English

• In language learners
  – Classroom routines e.g. *what’s your name?*
  – Rote-learned expressions e.g. *where’s the station?*
Psycholinguistic definition

“a sequence, continuous or discontinuous, of words or other elements, which is, or appears to be, prefabricated: that is, stored and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar.”

Wray, 2002: 9

“a multimorphemic unit memorised and recalled as a whole, rather than generated from individual items based on linguistic rules”

Myles et al. 1998
FS in the context of SLA

• Two different research agendas:
  1. Understanding how ‘external’ FS are learned: why are idioms difficult for learners? why do they seem to often use language which doesn’t sound quite ‘natural’ i.e. idiomatic?
  2. Understanding how learners process and analyse the input they are exposed to; what is the status of these rote-learned chunks in their interlanguage grammar?

In this paper, we are interested in 2. More specifically....
Research agenda

• Learner internal psycholinguistic definition:
  – What is processed as a single unit in learners (it might/might not be the same as in NS)
  – This processing advantage might be because the multiword unit is stored as a whole in the learner's lexicon or because it is highly automatised

• What is the role of FS in learner development, e.g. communicative crutches: processing shortcuts; planning tools?
  – In beginners
  – In advanced learners

• How do FS develop over time and how does their role change?

• What is their relationship with the development of other aspects of the linguistic system, e.g. grammar, lexis, fluency?
Identification issues

- Difficult: how can you tell a learner has processed a sequence as one unit or not?

- Relatively easy in beginners: FS are very different from the rest of a learner’s productions

- Much more difficult in advanced learners
Identification issues in beginners

1. Greater length and **complexity** of sequence compared with other learner output; usually well-formed (Weinert, 1995: 182)

   (1) *quel âge as-tu?*
   what age have you? (*wh*-fronting and verb-subject inversion) = how old are you?
   (Intended meaning: “how old is your brother?”)

   (2) *Il âge frère?*
   he age brother?
   (Intended meaning: “how old is his brother?”)

   Pupil 02, R4; produced by the same learner during the same elicitation task

   (all examples taken from study 1)
2. Phonological coherence, that is, **fluent**, non-hesitant encoding without a break in the intonation contour (Weinert 1995, p. 182)

(3) *quelle est la date de ton anniversaire?*  
what is the date of your birthday?  
= when is her birthday?

(4) *est ... elle bon anniversaire?*  
Is ... she happy birthday?  
= when is her birthday?  
(Pupil 02, R4; produced by the same learner during the same elicitation task)
3. They tend to be used inappropriately (syntactically, semantically, pragmatically), as numerous examples of overextensions clearly show:

(5) *mon petit garçon euh où habites-tu?
    my little boy umm where do you live?
    = where does your little boy live?
4. They are generally used in the same form, with no parts substitutable, that is, learners are not able to change any elements in the sequence, such as the second person reference in:

(6) *as-tu des frères ou des soeurs?*

have you any brothers or sisters?

= does she have any brothers or sisters?

There is no evidence elsewhere in the data of the *as-tu* construction being used in another context.
5. They tend to appear to be well-formed, and to be **grammatically advanced** compared to the rest of the learner’s language (i.e., unrelated to productive patterns in the learner’s speech):

(7) *comment t’appelles-tu?*  
what are you called?

(8) *euh une nom?*  
umm a name?

(Pupil 45, R5; same learner during the same task; intended meaning for both: *what is his name?*)
Identification issues in advanced learners

• Not the same as in beginners, i.e. in terms of...
  – Complexity
  – Fluency
  – Inappropriateness
  – Non-substitutability
  – Grammatically advanced/correct

• Advanced learners’ productions are already (usually!) fluent, complex, appropriate, varied and grammatical

• Crucial issue in advanced learners: evidence of a processing advantage if FS are processed as a single unit

• Operationalisation: how can you tell in an advanced learner (or a native speaker for that matter) that something is produced as a whole?
Identification method (study 2)

1. **Necessary** criterion: phonological coherence
   Fluent pronunciation of the multiword sequence, i.e.:
   - No filled or unfilled pauses longer than 0.2 second
   - No syllable lengthening
   - No repetition or retracing
   e.g. *it’s raining/pause/cats and/pause/dogs* is not a FS

2. **Necessary** additional presence of one typical criterion showing the unity of the sequence:
   - Grammatical or semantic irregularity, e.g. *dans le soir* (direct translation of *in the evening* – French NS would say *le soir*)
   - Holistic form-meaning/function mapping e.g. *by and large*
   - Likely presence of the sequence in input through instruction, e.g. *je suis (pas) d’accord* – I (dis)agree.

3. **Learner internal frequency** (not a necessary criterion)
   Graded criterion strengthening the case for formulaicity: preferred usage for a particular learner

Cordier 2013
The studies

• Early learners study datasets
  – Beginners
    • Longitudinal over 2 years; 60 learners (analysis here on a subset of 16 learners); years 7, 8, 9 (11-14 years old); up to 13 oral tasks per learner (all one-to-one with researcher or child-child pair work)

  – Post-beginners
    • Cross-sectional; 20 in each of years 9, 10, 11 (14-16 years old); 4 oral tasks (1:1 with researcher; 3 repeated from beginners study)

– Details on www.flloc.soton.ac.uk
Advanced learners’ study (Cordier 2013)

- 5 English university learners of French (similar educational background and level in French)

- Longitudinal study pre/post year abroad

- 2 data collection points:
  - End of 2nd year at university (immediately prior to compulsory year abroad in France)
  - During 7th month of stay in France

- Post year abroad questionnaire (about use of French during year abroad)

- 5 oral tasks repeated at time 1 and 2 (10 months apart): general 1:1 interview with researcher; picture based story telling task; 3 discussion tasks in pairs (approx 1 hr of data per learner each time)
Use of FS - Beginners

• The role of FS in the development of interrogative constructions

• All complex structures involving wh-fronting and subject verb inversion, well beyond the syntactic means of beginner learners

• Interrogative chunks tracked
  – *Où habites-tu?*
    where live you (where do you live?)
  – *Comment t’appelles-tu?*
    how yourself call you (what’s your name?)
  – *Quel âge as-tu?*
    what age have you (how old are you?)
  – *Quelle est la date de ton anniversaire?*
    what is the date of your birthday
Research questions

• how do these chunks, all present in all learners at the beginning of data collection, develop over time?
• How do interrogative constructions when learners have no chunks available develop over time?
• What is the link between the two?

• Asking ‘what’s his/her name?’
• Chunk: *comment t’appelles-tu?* 2nd person reference (what’s your name)
• How do learners ask ‘what’s his/her name?’
5 developmental stages for 3\textsuperscript{rd} person naming

1. Chunk over-extended: *comment t’appelles-tu?* (e.g. pointing at the person on the photo)
2. Chunk over-extended, but lexical NP tagged on: *comment t’appelles-tu le garçon?*
3. Chunk starting to break down: *comment t’appelle (la fille)?*
4. Further breaking down: comment s’appelle (un garçon)?
5. 3\textsuperscript{rd} person pronoun used (very rare): comment s’appelle-t-il?
Development of interrogatives when no chunk is available

1. Verbless stage
   *Je grand maison?* (I big house)

2. Non-finite verb stage
   *La mère regarder le magasin?* (the mother look – non-finite – the shop)

3. Finite verb stage
   *La mère regarde livre?* (the mother looks book)

This is in keeping with the development of the productive grammar generally
<table>
<thead>
<tr>
<th></th>
<th>Year 7 2\textsuperscript{nd} term</th>
<th>Year 7 3\textsuperscript{rd} term</th>
<th>Year 8 1\textsuperscript{st} term</th>
<th>Year 8 2\textsuperscript{nd} term</th>
<th>Year 8 3\textsuperscript{rd} term</th>
<th>Year 9 1\textsuperscript{st} term</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interrogatives</td>
<td>41 (95.3%)</td>
<td>129 (83.8%)</td>
<td>53 (82.8%)</td>
<td>235 (87.4%)</td>
<td>287 (79.5%)</td>
<td>182 (81.3%)</td>
</tr>
<tr>
<td>without verbs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interrogatives</td>
<td>2 (4.7%)</td>
<td>25 (16.2%)</td>
<td>11 (17.2%)</td>
<td>34 (12.6%)</td>
<td>74 (20.5%)</td>
<td>42 (18.8%)</td>
</tr>
<tr>
<td>with verb</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>43 (100%)</td>
<td>154 (100%)</td>
<td>64 (100%)</td>
<td>269 (100%)</td>
<td>361 (100%)</td>
<td>224 (100%)</td>
</tr>
<tr>
<td></td>
<td>Year 7</td>
<td>Year 8</td>
<td>Year 11</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2\textsuperscript{nd} person chunk</td>
<td>18 (52.9%)</td>
<td>31 (39.2%)</td>
<td>27 (22.7%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3\textsuperscript{rd} person reference</td>
<td>16 (47.1%)</td>
<td>48 (60.8%)</td>
<td>92 (77.3%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>34 (100%)</td>
<td>79 (100%)</td>
<td>119 (100%)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Use of *comment t’appelles-tu* with 3\textsuperscript{rd} person reference
Relationship between chunks and generative grammar

At the same point in time, we find within individual learners both:

- Highly complex chunks, delivered relatively fluidly and accurately, e.g. *comment t’appelles-tu?*

- Very simple syntax outside chunks, e.g. *euh nom*? (pointing at picture)
Conclusions: role of FS in early stages

- All learners use FS, but some are much better at memorising them.

- The learners who have the largest store of FS keep using them (i.e. they do not discard them), but they are also the learners who are most active at ‘working’ on them to use their constituent parts elsewhere.

- FS give learners an entry into communication, before their interlanguage grammar can produce complex structures.

- FS provide a set of language samples which act as models for constructing their grammar.
The role of FS in advanced learners

• Identification problem: determining that a sequence has been processed holistically:
  the sequence presents a processing advantage and is retrieved faster than other strings of words, but also has some kind of unitary meaning

• Example of measures used:
  – Pauses (silent and filled; over 0.2 secs)
  – Identification of fluent runs (no interruptions/pauses)
  – Application of the other identification criteria

Using Praat software
Some measures of formulaicity used

• Total number of syllables
• Total number of syllables belonging to FS
  → % FS syllables / total number of syllables (%FS)

• Total number of runs
• Total number of runs containing FS
  → % FS runs / total number of runs (%RFS)
Some measures of fluency used

• Phonation/time ratio (% time spent speaking out of total task time)

• Mean length of runs (expressed in syllables)

• Speaking rate (no of syllables per sec, inc. pauses)

• Articulation rate (as above, exc. pauses)
Measure of lexical diversity

• TTR (type/token ratio) is not an accurate measure as it depends on length of text (the longer the text, the more words are repeated)

  The small cat was sleeping on the small mat

  9 tokens; 7 types

• D: measure which incorporates a mathematical formula taking account of this
Role and development of FS over time

- These measures of formulaicity, fluency and lexical diversity enable us to document:
  - Changes over time after an extensive period of immersion
  - Relationships (if any) between FS use, fluency and lexical diversity
Development of FS in advanced learners

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>SD (learner)</th>
<th>SD (t3.71ask)</th>
</tr>
</thead>
<tbody>
<tr>
<td>% FS Time 1</td>
<td>25.06</td>
<td>3.75</td>
<td>5.33</td>
</tr>
<tr>
<td>% FS Time 2</td>
<td>30.48</td>
<td>3.32</td>
<td>3.71</td>
</tr>
<tr>
<td>% RFS Time 1</td>
<td>31.86</td>
<td>7.04</td>
<td>8.21</td>
</tr>
<tr>
<td>% RFS Time 2</td>
<td>46.1</td>
<td>10.41</td>
<td>7.79</td>
</tr>
</tbody>
</table>

%FS: % formulaic syllables / total syllables
%RFS: % runs containing FS

Differences statistically significant
FS: $F_s(1,4)=79.94$, $p=0.001$; $F_t(1,4)=26.36$, $p=0.007$
RFS: $F_s(1,4)=21.28$, $p=0.0101$; $F_t(1,4)=463.82$, $p<0.0001$

No significant differences across learners and across tasks (except for one learner who started from a lower base)
Increase in percentage of FS per learner between time 1 and time 2

<table>
<thead>
<tr>
<th></th>
<th>T1</th>
<th>T2</th>
<th>+%</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Iris</td>
<td>25.7</td>
<td>33</td>
<td>+7.3</td>
<td>0.13</td>
</tr>
<tr>
<td>Lily</td>
<td>24.56</td>
<td>29.66</td>
<td>+5.1</td>
<td>0.01*</td>
</tr>
<tr>
<td>Lola</td>
<td>26.78</td>
<td>31.98</td>
<td>+5.2</td>
<td>0.18</td>
</tr>
<tr>
<td>Rose</td>
<td>19.1</td>
<td>25.02</td>
<td>+5.92</td>
<td>0.06</td>
</tr>
<tr>
<td>Sally</td>
<td>29.18</td>
<td>32.74</td>
<td>+3.56</td>
<td>0.19</td>
</tr>
</tbody>
</table>
Increase in percentage of FS per learner between time 1 and time 2
### Average no of formulaic syllables per run

<table>
<thead>
<tr>
<th></th>
<th>T1</th>
<th>T2</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Iris</td>
<td>1.13</td>
<td>2.3</td>
<td>0.0000*</td>
</tr>
<tr>
<td>Lily</td>
<td>0.97</td>
<td>1.88</td>
<td>0.0026*</td>
</tr>
<tr>
<td>Lola</td>
<td>1.29</td>
<td>1.96</td>
<td>0.0108*</td>
</tr>
<tr>
<td>Rose</td>
<td>0.67</td>
<td>0.99</td>
<td>0.0684</td>
</tr>
<tr>
<td>Sally</td>
<td>1.53</td>
<td>2.27</td>
<td>0.0060*</td>
</tr>
</tbody>
</table>
Average no of formulaic syllables per run

![Bar chart showing average formulaic syllables per run for different names (Iris, Lily, Lola, Rose, Sally). Each name has two bars, one for T1 and one for T2. The y-axis represents the number of syllables, ranging from 0 to 2.5.]
Summary development of FS during year abroad

• All learners increase their FS use significantly on at least some of the measures:
  – More FS
  – Longer FS
Lexical diversity at time 1 and time 2

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>SD (subjects)</th>
<th>Min (subjects)</th>
<th>Max (subjects)</th>
</tr>
</thead>
<tbody>
<tr>
<td>D time 1</td>
<td>58.89</td>
<td>5.24</td>
<td>53.76 (I)</td>
<td>65.02 (S)</td>
</tr>
<tr>
<td>D time 2</td>
<td>71.46</td>
<td>11.54</td>
<td>54.10 (R)</td>
<td>86.20 (S)</td>
</tr>
</tbody>
</table>

Increase significant: \( F_s(1,4) = 9.91, p=0.035; F_t(1,4)=30.28, p=0.005 \)
Lexical diversity per learner at Time 1 and Time 2
Development of FS use in relation to development of lexical diversity

All learners develop the two in parallel except Rose
Correlations between FS use and lexical diversity

<table>
<thead>
<tr>
<th></th>
<th>D 1</th>
<th>D 2</th>
<th>D change</th>
</tr>
</thead>
<tbody>
<tr>
<td>%FS 1</td>
<td>0.292</td>
<td>0.746</td>
<td>0.619</td>
</tr>
<tr>
<td></td>
<td>p=0.200</td>
<td>p=0.000*</td>
<td>p=0.003*</td>
</tr>
<tr>
<td>%FS 2</td>
<td>0.122</td>
<td>0.422</td>
<td>0.375</td>
</tr>
<tr>
<td></td>
<td>p=0.597</td>
<td>p=0.057</td>
<td>p=0.094</td>
</tr>
<tr>
<td>%FS diff</td>
<td>-0.181</td>
<td>-0.374</td>
<td>-0.292</td>
</tr>
<tr>
<td></td>
<td>p=0.433</td>
<td>p=0.095</td>
<td>p=0.200</td>
</tr>
</tbody>
</table>

- Significant correlation between FS use at time 1 and lexical diversity at time 2
- Significant correlation between FS use at time 1 and the size of lexical development between time 1 and time 2
Summary relationship FS use and lexical diversity

• The more a learner is formulaic at time 1, the more likely s/he is to be lexically diverse at time 2

• The automatisation process involved in FS is crucial for freeing attentional resources to learn more vocabulary/grammar
Fluency development: amount of speech

<table>
<thead>
<tr>
<th></th>
<th>Iris</th>
<th>Lily</th>
<th>Lola</th>
<th>Rose</th>
<th>Sally</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>T1</strong></td>
<td>3877</td>
<td>3316</td>
<td>3631</td>
<td>2974</td>
<td>5689</td>
</tr>
<tr>
<td><strong>T2</strong></td>
<td>10020</td>
<td>6437</td>
<td>7343</td>
<td>4191</td>
<td>7028</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>13896</td>
<td>9753</td>
<td>10974</td>
<td>7165</td>
<td>12717</td>
</tr>
</tbody>
</table>

No of tokens used per learner at time 1 and time 2 (same tasks):
Learners speak considerably more at time 2
Fluency measures: development over time

<table>
<thead>
<tr>
<th>Measure</th>
<th>Time (df1,4)</th>
<th>Time x Subjects (df4,16)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PTR</td>
<td>26.29, p = 0.007*</td>
<td>10.80, p &lt; 0.001*</td>
</tr>
<tr>
<td>MLR</td>
<td>250.60, p &lt; 0.001*</td>
<td>9.84, p &lt; 0.001*</td>
</tr>
<tr>
<td>SR</td>
<td>384.00, p &lt; 0.001*</td>
<td>13.70, p &lt; 0.001*</td>
</tr>
<tr>
<td>AR</td>
<td>188.52, p &lt; 0.001*</td>
<td>14.25, p &lt; 0.001*</td>
</tr>
</tbody>
</table>

PTR: phonation/time ratio; MLR: Mean length of runs (syllables); SR: speaking rate; AR: articulation rate
Increase in PTR per learner between time 1 and time 2
Increase in MLR per learner between time 1 and time 2
Increase in SR per learner between time 1 and time 2
Increase in AR per learner between time 1 and time 2
Summary: development of fluency

- Learners speak faster, with less hesitations and pauses

- Large improvement in all learners except Rose
Close relationship between fluency (as measured by speaking rate) and FS use.
Summary – advanced learners

• FS do not disappear when lexical diversity increases: they become more varied and sophisticated;

• FS use and lexical diversity correlate: the more learners use FS, the more diverse their lexis is, and the more they improve on both counts over time

• FS use and fluency correlate: the more learners use FS, the more fluent they are overall

• Advanced learners still rely on the FS they learned in beginner classrooms
Conclusion

At both ends of the continuum:
• FS are used to facilitate speech production, but in different ways
  – Early learners use them as communicative crutches and as a model for hypothesis forming/testing
  – Advanced learners use them as processing shortcuts and to buy planning time. They also use them to free up attentional resources to pay attention to novel aspects of the language they are learning