Information structural transfer in advanced Dutch EFL writing
A cross-linguistic longitudinal study*

Sanne van Vuuren
Radboud University Nijmegen

This article presents a case study on the role of L1 transfer of language-specific features of information structure in very advanced L2 learners. Cross-linguistic differences in the information status of clause-initial position in a V2 language like Dutch compared to an SVO language like English are hypothesized to result in overuse of clause-initial adverbials in the writing of advanced Dutch learners of English. This hypothesis was tested by evaluating advanced Dutch EFL learners’ use of clause-initial adverbials in a syntactically annotated longitudinal corpus of student writing, compared to a native reference corpus. Results indicate that Dutch EFL learners overuse clause-initial adverbials of place as well as addition adverbials that refer back to an antecedent in the directly preceding discourse. Although there is a clear development in the direction of native writing, transfer of information structural features of Dutch can still be observed even after three years of extended academic exposure.

Keywords: second language acquisition, information structure, L1 transfer, advanced learners, EFL, clause-initial position

1. Introduction

Dutch students of English Language and Culture are expected to reach a near-native level of proficiency, and indeed their writing tends to be relatively error-free (Springer 2012; de Haan & van der Haagen 2012). Then why is it that even at these advanced stages of acquisition their writing is often considered to be recognizably Dutch? As most language acquisition research has focused on earlier stages of L2 acquisition, relatively little is known about advanced learners, but it is clear that, as Carroll and Lambert (2003: 270) have noted, “the learning problem at advanced stages of learning is not one of linguistic form”. Rather, it appears that advanced
learners differ from native speakers in (1) the frequency with which they use lexico-grammatical devices and syntactic structures available in the language (see for example Springer 2012; de Haan & van der Haagen 2012) and (2) in the application of language-specific principles of information structure (Bohnacker and Rosén 2008; Callies 2009; Verheijen, Los & de Haan 2013). Callies, for example, notes that

> [e]ven at advanced stages of L2 acquisition, in which ILs can be considered near-native in many respects, some core principles of information structure typical of those found in the native languages of the learners are retained, and divergences can be attributed to fundamental principles of organization underlying information structure. (2009: 104)

Similarly, following Verheijen et al. (2013), we hypothesize that the appearance of non-nativeness of texts written by advanced Dutch learners of English is largely due to the frequency with which these learners use certain types of clause-initial adverbials in their L2. This in turn might be attributed to transfer caused by an interaction of syntactic and information structural differences in the use of clause-initial constituents between Dutch, a verb-second language with a multifunctional clause-initial position, and English, which has a more rigid SVO structure (Los 2009). It is against this background that this article aims to quantify and evaluate advanced Dutch EFL learners’ use of clause-initial adverbials in a syntactically annotated longitudinal corpus of student writing, compared to a native reference corpus. In doing so, we hope to answer the following questions:

1. How do advanced Dutch EFL learners differ from native speakers in the frequency with which they use clause-initial adverbials and in the way they use these to provide a link to the preceding discourse?
2. Can a development be observed in advanced Dutch EFL learners’ use of clause-initial adverbials in the direction of native writing?

Apart from providing further insight into the role of information structural transfer in second language learning, the answers to these questions may have implications for language teaching at advanced stages of acquisition.

2. Comparative pragmatics of clause-initial position in Dutch and English

Clause-initial position has a key role in linking a clause to the preceding discourse and in providing a background against which the message in the remainder of the clause is to be interpreted (Virtanen 2004). In Dutch, a verb-second language, clause-initial position is both syntactically and information-structurally
“multifunctional”, in the sense that it may be occupied by either a subject (1a), an object (1b) or an adverbial (1c) and, depending on context, these constituents may be either marked or unmarked (Los 2009):

(1) a. Andy Cole heeft hier zijn eerste hattrick gescoord.
Andy Cole has here his first hat-trick scored

His first hat-trick has Andy Cole here scored

c. Hier heeft Andy Cole zijn eerste hattrick gescoord.
Here has Andy Cole his first hat-trick scored

(Adapted from Hannay & Keizer 1993: 68)

In Dutch, clause-initial position commonly hosts what Los and Dreschler refer to as ‘local anchors’, adverbials which serve to link the sentence they occur in to the immediately preceding discourse (2012: 859). This tendency to link sentences together by means of (unmarked) clause-initial adverbials is shared by other V2 languages such as Norwegian (Hasselgård 2009) and German (Kirkwood 1969; Carroll, Murcia-Serra, Watorek & Bendiscioli 2000; Bohnacker & Rosén 2008).1 English, on the other hand, has a more rigid SVO structure in which the use of clause-initial position is restricted and the subject has an important linking function (Carroll et al. 2000; Los 2009). Dreschler & Hebing (2010: 64), for example, find that 77% of English sentences start with a subject, while only 23% start with other first constituents (against 54% subjects and 46% other first constituents for Dutch). It is clearly not ungrammatical in English for adverbials to occur in front of the subject, but, as Los (2009: 26) argues, “[t]he very fact that this presubject position does not need to be filled, unlike the first position in a verb-second system, makes it likely that it has acquired a special, marked position”.

With 64 occurrences per 1000 words (against 14 for initial position and 20 for medial position), in English “final position is by far the most common position for adverbials” (Biber, Johannsson, Leech, Conrad & Finegan 1999: 772). This can be accounted for by the preference for this position that has been found for the large and varied class of circumstance adverbials, most of which are realized as prepositional phrases (Biber et al. 1999: 772, 807). Interestingly, Biber et al. note that, especially in written registers, initial prepositional phrases may “have a cohesive function, with the prepositional phrase using some information given in the previous discourse as the starting point for the next sentence” (1999: 809). This is exactly what we find in Dutch and so apparently contradicts the idea that there are any fundamental information-structural differences between both languages. However, differences between Dutch and English in the frequency with which such phrases are used and the contexts in which they commonly appear
point to information structural constraints associated with the use of clause-initial position in English sentences.

Translation manuals, such as Lemmens and Parr (1995:92), for example, generally warn against using adverbials at the start of an English sentence and suggest that the noun following the preposition in the Dutch adverbial in (2a) should be used as a subject instead (see (2b) below), in order to avoid giving the adverbial, which was unmarked in Dutch, too much emphasis in its English translation, which is what happens in (2c):

(2) a. Met de geldautomaat heeft u altijd toegang tot uw rekening.
    With the cash dispenser have you always access to your bank account
b. The cash dispenser gives you constant access to your bank account.
c. With the cash dispenser you always have access to your bank account.
    (adapted from Lemmens & Parr 1995:92)

Also consider the difference between the following sentences:

(3) a. In Nederland zijn de meeste scholen openbaar.
    In the Netherlands are the most schools public
b. Education is usually public in the Netherlands.
c. In the Netherlands, education is usually public. (In Brazil, on the other hand, most middle-class children attend private schools.)

Compared to the clause-final adverbial in (3b), the clause-initial adverbial in (3c) is much more marked and is likely to be interpreted as contrastive. While Dutch circumstance adverbials, such as the instrument adverbial in (2a) and the place adverbial in (3a), are typically used in clause-initial position to function as unmarked local anchors, in English the only reason to move them to clause-initial position would be to give them extra prominence or contrastive focus and it is the subject of the sentence which is the prime candidate for establishing an unmarked link to the preceding discourse.

Dutch students of English apparently lack awareness of these subtle cross-linguistic differences, which manifests itself in their typical overuse of certain types of clause-initial adverbials, particularly those that are meant to function as unmarked local anchors. Take, for example, the following sentences from a text by a Dutch student writer:

(4) Besides this, I have also been active in my student society where I was responsible for the finance of two committees involved in the organization of our lustrum.

(5) Due to my internship I realized that students are already capable of implementing their knowledge into business.
(6) *Therefore*, I started the Student Consultancy Group which brings top students and companies together.

While the Dutch translations of the adverbials in italics could function as neutral discourse links, in English their position in front of the subject is much more marked. A native speaker might have left out ‘besides this’ from sentence (4). ‘Due to my internship’ in (5) implies contrast due to its evocation of alternative sources of realization and might have been replaced by ‘My internship made me realize’, while ‘Therefore’ in (6) could have been replaced by a reverse *wh*-cleft: ‘This is why…’. In each case, it would be more natural to start the sentence with a subject rather than an adverbial.

3. Method

3.1 Corpus design

We evaluated the use of clause-initial adverbials in a longitudinal corpus of 899 student essays written by Dutch students of English Language and Culture between their first and third year at university. These were collected between 2008 and 2012 at Radboud University Nijmegen as part of the LONGDALE project, a European project aimed at compiling longitudinal corpora of texts written by advanced EFL learners with different language backgrounds (Granger 2009). As previous studies have provided clear evidence that the frequency with which different categories of adverbials are used in clause-initial position largely depends on genre (see for example Biber et al. 1999 and Bohnacker & Rosén 2008), we distinguished between two types of writing included in the Dutch component of LONGDALE: (1) 440 essays on various aspects of British or American literature and culture with a total word count of 481,956 and (2) 459 timed, argumentative in-class writing assignments on subjects such as ‘the pros and cons of an obligatory stay abroad for BA students’ or ‘the need for conservation at Radboud University’. As the in-class assignments, which had to be completed within 30 minutes, are considerably shorter than the essays, the total word count for the in-class assignments adds up to 146,481.

In order to allow an initial comparison to the level of academic writing that students of English Language and Culture are ultimately aiming to achieve, Philip Springer’s VU Native Speaker Published Research Article Corpus, consisting of 11 linguistics articles (79,121 words) and 11 literature articles (70,890 words), was used as a reference corpus (Springer 2012).
3.2 Procedure

Syntactic annotation was added to the corpora using the Stanford Parser (Klein & Manning 2003) after which adverbial phrases occurring in clause-initial position in declarative main clauses were filtered out with Corpus Studio (Komen 2011). This procedure resulted in a database of 8774 clause-initial adverbials for the learner corpus (13.96 per 1000 words) and 1891 for the reference corpus (12.61 per 1000 words). These adverbials were then categorized according to their function label (e.g. ‘instrument’, ‘addition’, ‘place’) as well as their discourse status (‘identity’, ‘inferred’, ‘assumed’, ‘new’ or ‘inert’) and distance to their antecedent (-1, -2, etc.). If the adverbials were part of a quote, they were not taken into account.

3.3 Classification of adverbials

3.3.1 Semantic categories

Because some types of adverbials can occur in clause-initial position more easily than others, labels were added to distinguish between different semantic categories. The labels that were used in the annotation process are based on Biber et al. (1999: 762–892), apart from the category of ‘domain’, which is based on Mittwoch, Huddleston & Collins (2002: 765–766) and which serves to distinguish those adverbials which “restrict the domain to which the rest of the clause applies” (765), such as ‘with regard to’, ‘on the subject of’ and ‘as for’. Table 1 presents an overview of the categories that were used.

<table>
<thead>
<tr>
<th>Semantic Category</th>
<th>Subcategory 1</th>
<th>Subcategory 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>accompaniment</td>
<td>doubt and certainty</td>
<td>recipient</td>
</tr>
<tr>
<td>actuality and reality</td>
<td>enumeration</td>
<td>restriction</td>
</tr>
<tr>
<td>addition</td>
<td>exemplification</td>
<td>result/inference</td>
</tr>
<tr>
<td>apposition</td>
<td>extent/degree</td>
<td>style</td>
</tr>
<tr>
<td>attitude</td>
<td>instrument</td>
<td>summation</td>
</tr>
<tr>
<td>cause/reason</td>
<td>limitation</td>
<td>time</td>
</tr>
<tr>
<td>comparison</td>
<td>manner</td>
<td>transition</td>
</tr>
<tr>
<td>condition</td>
<td>means</td>
<td>viewpoint or perspective</td>
</tr>
<tr>
<td>contrast/ concession</td>
<td>place</td>
<td></td>
</tr>
<tr>
<td>domain</td>
<td>purpose</td>
<td></td>
</tr>
</tbody>
</table>

If an adverbial phrase contained a relative clause, the phrase received an extra label, so that ‘due to the corruption that was going on’, for example, would get the label ‘cause/reason — incl rel clause’. This paper focuses on the categories of
3.3.2 Referential state categories: ‘the pentaset’

Apart from semantic function category labels we also added information about each adverbial’s referential state category. This was done using ‘the pentaset’, a set of five referential state primitives which can be combined with syntactic information to derive the information structure of a sentence (Komen 2013: 122). These referential state primitives are (1) identity, for those adverbials containing an NP that refers to the same entity as its antecedent; (2) inferred, for adverbials containing an NP that can be inferred from its antecedent; (3) assumed, for adverbials containing an NP that refers to an entity outside the text; (4) new, for adverbials containing an NP that is newly introduced, i.e. without an antecedent; and (5) inert, for adverbials containing an NP that cannot, in fact, be referred back to at all, such as the word ‘addition’ in the phrase ‘in addition’ (Komen 2013). For each adverbial with a textual antecedent, that is to say, those adverbials that received the label ‘identity’ or ‘inferred’, we also indicated the distance to its antecedent. If the adverbial had an identity relation to its antecedent and the antecedent was located in the directly preceding sentence, it received the label ‘identity, -1’; if the adverbial was three sentences removed from its antecedent, it received the label ‘identity, -3’; and so on.

4. Results and discussion

4.1 Text types

The data show a clear distinction between the essays and the in-class assignments. While the essays typically use many place adverbials in clause-initial position to establish a link with the preceding context by means of references to the text that is being discussed, addition adverbials predominate in the in-class assignments. Both categories will be discussed in turn.

4.2 Place

While Dutch place adverbials commonly occur clause-initially to provide a link to the preceding context, previous research by Biber et al. has shown that in English only 5% of place adverbials occur in clause-initial position, another 5% in medial position, while an overwhelming 90% occur in clause-final position (Biber et al. 1999: 802).3 This difference between Dutch and English is reflected in a steady
decline in the use of clause-initial place adverbials in the LONGDALE corpus between year 1 and 3 (Figure 1). While a decline can be observed for both text types, they have widely different starting points. The essays start out at 2.63 clause-initial place adverbials per 1000 words in year 1, more than three times as many as for the in-class assignments, which start out at 0.84 per 1000 words. This can be explained by the high number of references to the text that is being discussed in literature essays, such as *in the poem* or *in chapter three*.

Figure 2 visualizes the decrease between year 1 and year 3 in the number of clause-initial place adverbials that function as local anchors because they link back to the directly preceding context by means of an identity link. The essays go from

---

**Figure 1.** Clause-initial place adverbials per 1000 words in LONGDALE

**Figure 2.** Clause-initial place adverbials with an identity link to -1 per 1000 words in LONGDALE
Information structural transfer in advanced Dutch EFL writing

0.64 in year 1 to 0.37 in year 3, with a slight dip in year 2, but remain well above the level of the literature articles in the reference corpus, which use only 0.24 clause-initial place adverbials with an identity link to -1 per 1000 words (Figure 3). The frequency of clause-initial place adverbials used in the in-class assignments, on the other hand, both overall and those that link back to -1, is considerably lower than the frequencies found in the reference corpus. This is probably due to the fact that the subjects which students were asked to write about did not lend themselves very well to text structuring by means of chains of place adverbials. Differences between the two types of texts included in the reference corpus, finally, also point towards a strong effect of genre. A closer look at the use of place adverbials in the linguistics articles reveals that the high number of adverbials of this type can be accounted for by the frequent use of phrases that refer to the text itself, such as below, in this article, in this section, in table two, in examples 9 and 10, etc., which appear to be less common in literature articles.

4.3 Addition

Although the number of adverbials in clause-initial position in general might be expected to decline due to decreasing interference from Dutch with its preference for local anchors, Figure 4 shows that the use of clause-initial adverbials that fall into the category of addition actually increases between year 1 and 3, only slightly for the essays, but more considerably for the in-class assignments, which start out at 1.43 clause-initial addition adverbials per 1000 words and end up at 1.71 by year three. This might be attributed to the fact that in English, too, addition adverbials, which fall into the class of linking adverbials, have a preference for initial position,
especially in academic prose (Biber et al. 1999: 891), and students are explicitly taught to structure their writing and to make sure they link their sentences together by means of cohesive devices.

While there is an increase in the number of clause-initial adverbials in the category of addition overall, a different picture emerges from a sub-classification according to referential state. Figure 5 shows that there is a steady decrease in the number of addition adverbials with an identity link to the directly preceding context, such as in addition to this or on top of that. Again, this development is more marked for the in-class assignments than for the essays.

Interestingly, the number of addition adverbials in initial position in the reference corpus is in fact lower than the starting level for the category of addition

Figure 4. Clause-initial addition adverbials per 1000 words in LONGDALE

Figure 5. Clause-initial addition adverbials with an identity link to -1 per 1000 words in LONGDALE
adverbials in year one in the learner corpus. Figure 6 shows that the native speaker literature articles (NS lit) and the native speaker linguistics articles (NS ling) use 0.39 and 0.76 addition adverbials per 1000 words respectively, whereas the essays in the learner corpus (Figure 4), which are similar in genre to the NS literature articles and might therefore be expected to be comparable in style, start out at 0.77 and increase up to 0.82 by year three. The number of addition adverbials used in the in-class assignments is even higher, starting out at 1.43 and increasing up to 1.71 by year three. Although the reference corpus does not contain any writing that can be said to belong to the same genre as the timed, argumentative in-class writing assignments included in the learner corpus, it is clear that the number of initial addition adverbials in the in-class assignments is not just higher than that of the essays in the learner corpus but also higher than that of either category in the reference corpus.

What is even more striking is the fact that the category of ‘addition, -1’ is virtually non-existent in the reference corpus (Figure 6), which means that while there is a downward trend for the learners of English in the direction of native writing in this respect (see Figure 2), by year 3 there is still a gap between the number of addition adverbials with an identity link to -1 in the essays and in-class assignments in the learner corpus (0.04 and 0.14 per 1000 words respectively) and the number of addition adverbials in both the linguistics and the literature articles in the reference corpus (0.01 per 1000 words for both groups).
5. Conclusion

The case study presented here shows that differences between Dutch and English in the information status of clause-initial position do lead advanced Dutch EFL learners to overuse categories of adverbials, such as place or addition adverbials, which in Dutch are commonly used in clause-initial position to provide a link to the preceding discourse. However, there are interesting differences between these two categories. Place adverbials, like most circumstance adverbials, only rarely occur in clause-initial position in English sentences, of which students seem to become increasingly aware between their first and third year at university. The number of place adverbials that refers back to the directly preceding context declines at a similar pace. Addition adverbials, on the other hand, belong to the class of linking adverbials, which in fact have a preference for initial position even in English. As a result, students seem to be less aware of the infelicity of the use of addition adverbials in clause-initial position in their L2, judging from the increase in this category between year 1 and year 3. Interestingly, a development in the direction of native writing can only be observed when addition adverbials are subcategorized according to referential status, which brings to light a steady decline in the number of addition adverbials that have an identity relation with an antecedent in the directly preceding context. Still, it is in this subcategory that the biggest contrast between native writing and the writing of advanced Dutch EFL writers can be observed, even by the end of year 3, as addition adverbials that refer back to an antecedent in the directly preceding context are virtually non-existent in the reference corpus.

These results are of course particularly relevant for those who, like our students of English Language and Culture, are not just EFL users but future EFL professionals (de Haan & van der Haagen 2012) and therefore need to reach beyond grammatical correctness to a near-native level of proficiency. Many researchers have observed that exposure to the target language alone is not sufficient to learn to recognize subtle cross-linguistic differences (Norris & Ortega 2000; Hinkel 2003). If non-native writers wish to acquire a near-native level of proficiency, they will need teaching materials focusing on those areas which distinguish advanced learners from native speakers. At present, most textbooks are not informed by corpus research. The few exceptions to the rule are based on native speaker corpora and not specifically aimed at non-native writers (Gilquin, Granger & Paquot 2007). Further corpus research into transfer of language-specific principles of information structure should therefore serve as an empirical basis for the development of L1-specific EFL teaching materials.
Notes

* I would like to thank my supervisors Ans van Kemenade, Pieter de Haan and Bettelou Los as well as my colleague Erwin Komen for their valuable contribution to this project. Many thanks also to the audience at the 2013 TIN-dag, the contributors to Radboud University’s Language in Transition Stages workshops and two anonymous reviewers for their helpful comments.

1. Interestingly, speakers of Swedish prefer to start their sentences with subjects or expletives in spite of the fact that Swedish has the same V2 system in which clause-initial position can be occupied by almost any type of constituent (Bohnacker & Rosén 2008: 517). This suggests that frequent use of local anchors is related not just to the availability of a multifunctional first position in V2 languages, but also requires a “system of deictic referencing [which is] sufficiently articulated” (Los 2009: 37).

2. While there is not a great difference between the total number of clause-initial adverbials in the two corpora overall, a subclassification according to semantic category and referential state, as outlined below, reveals interesting differences between the Dutch EFL learners and the native speakers used as a reference.

3. As the focus of the present study is on clause-initial adverbials, rather than the distribution of adverbials across initial, medial and final position, we cannot directly compare these results to our own data. The Longman Corpus used by Biber et al. therefore does not serve as reference material. However, it does serve to illustrate that English place adverbials have a clear preference for final position.

References


Author’s address

Sanne van Vuuren
Department of English Language and Culture/CLS
Radboud University Nijmegen
Postbus 9103
6500 HD Nijmegen, The Netherlands
s.v.vuuren@let.ru.nl