Data-Driven Learning: From Classroom Scaffolding to Sustainable Practices

Fanny Meunier
Université catholique de Louvain, Belgique

Abstract This article calls for more constructive alignment in the DDL (Data-Driven Learning) pratices currently carried out in instructed settings. After a short description of constructive alignment, this study addresses some of the reasons behind the lack of alignment that, in turn, often leads to a lack of uptake and sustainable practices in DDL. Concrete examples of how courses including DDL activities could be better aligned are presented. Solutions to increase sustainability are also proposed; they include the need for terminological simplification and a revision of some language assessment practices. This article also stresses the importance of both actively listening to teachers' and learners'needs and concretely supporting teachers' efforts to embrace DDL.

Keywords Alignment. Sustainable practices. Terminology. Assessment. Teacher support.

Summary 1 Introduction. – 2 The Need for More Constructive Alignment in DDL: Focus on Formal Instructed Settings. – 3 Reasons for the Current Lack of Sustainable Practices in DDL. – 3.1 A Plea for Terminological Simplification. – 3.2 Constructive Alignment Includes Assessment. – 4 Conclusion.

1   Introduction

The term data-driven learning (DDL) was coined in 1991 by Johns and refers to the fact that learners and teachers can use and query/use corpus data for language awareness-raising tasks. Other articles in the present volume provide discussions and illustration of DDL but readers are also referred to Chambers (2010) for a detailed history of DDL and its practices. Vyatkina and Boulton (2017), Boulton (2017), and Boulton and Cobb (2017) contain in-depth discussions and illustrations of the potential of data-driven learning in language learning. As for recent publications specifically addressing the potential of DDL in the Italian context, I refer readers to Forti (2019), Corino and Onesti (2019).

In Meunier (2019) I describe DDL activities as a specific type of consciousness-raising tasks that exploit the affordances of corpus tools. Two of these affordances are the access to frequency data in a language (i.e. the frequent and typical words, expressions, structures of a given target language) and to the patterned nature of language, thanks to the visual setup and sorting facilities offered by concordancers, for instance. Native and learner corpora provide the data from which learners can discover language patterns for themselves; tools like concordancers make it possible to expose learners to input flooding and input enhancement, two concepts put forward by Sharwood Smith back in 1993 and which have been shown to be beneficial for language learning. As corpora are large collections of texts – which typically contain millions of words –, they often include numerous examples (input flooding) of the target form that teachers want to discuss with their learners. Corpus tools, for their part, can be used to easily retrieve those examples and organise them in specific ways so as to make their linguistic form more salient for learners (input enhancement). A quick look at left hand sorted concordances of the verb *prevent* will reveal that it is frequently followed by a *pronoun* and/or by the preposition *from*, itself followed by the gerundive form of a verb. Teachers can further enhance the output of concordances by adding extra typographical enhancement (bolding, colouring, etc.). In 1999, Ellis reviewed eight studies on input enhancement and tentatively concluded that input enhancement and flooding promoted the learning of the target form. As for Han, Park and Combs (2008, 69), they concluded on the basis of 21 studies that it was "compound enhancement" that benefited language learning, i.e. prior knowledge of the form and instructions to pay attention to the form. More recently, Forti (2019, ii) referred to recent meta-analyses indicating that DDL is "a generally effective approach in second language learning, worthy of being integrated in existing teaching and learning practices (Mizumoto, Chujo 2015; Boulton, Cobb 2017)". However, she also explained that these meta-analyses

"reveal that the effects of the approach vary considerably when taking into account a number of moderator variables, such as teaching context, proficiency level of the learners and type of study design investigating these effects" (2019, ii).[1] As for Corino and Onesti (2019, 2) they stated that good practices in DDL "align with current theories and practices of Second Language Acquisition, namely the constructivist and learner-centered approaches to language acquisition". The authors also added that the use of authentic language materials present in corpora is in line with communicative approaches to language learning and underpins the mandate "for the development in learners of metalinguistic knowledge and learner autonomy" (Godwin-Jones 2017).

The role of teachers in DDL activities is to act as facilitators in the development of metalinguistic knowledge by manipulating the concordances to scaffold the awareness raising activities (see Mishan, Timmis 2015, 81-2, or Corino, Onesti 2019 for concrete illustrations). By providing learners with DDL activities and gradually guiding them on how to best analyse them, teachers help learners become active constructors of knowledge instead of being passive recipients. Such activities require the mobilisation of higher level cognitive skills. Ideally, learners progressively become more proficient in corpus literacy and are expected to be able to use corpora and interpret the data independently when looking for some linguistic information needed to perform certain tasks (e.g. looking for frequent collocations of specific terms when writing a paper).

2 The Need for More Constructive Alignment in DDL: Focus on Formal Instructed Settings

Despite Corio and Onesti's statement that good practices in DDL "align with current theories and practices of Second Language Acquisition" and underpin "the mandate in contemporary communicative language instruction for the use of authentic language materials and for the development in learners of metalinguistic knowledge and learner autonomy" (2019, 2), very few DDL studies provide explanations and/or justifications on how the activities fit in the broader curricular context in which they are carried out. The DDL activities found in the literature are often presented in isolation and typically adopt the following pattern: what is DDL? why is it useful? what has been done with the pupils/students to focus on form x? to what extent

---

[1] For a lengthier discussion of the limitations and potential drawbacks of data-driven learning, also see Boulton 2017 and Forti 2019.

have these activities been useful for language learning?

In this paper I would like to argue for more constructive alignment in DDL as it constitutes an essential first step towards sustainable practices. Put simply, constructive alignment is an outcome-based approach related to both constructivist learning theory and instructional design literature. Biggs (1996, 347) explains that the "centrality of the learner's activities in creating meaning" found in constructive alignment came from constructivist theories, and the need for "alignment between the objectives of a course or unit and the targets for assessing student performance" from instructional design. Described as a backward type of design by Wiggins and Mc-Tighe (2005), a constructively aligned activity should:

– first, define the intended learning outcomes (viz. the content to be learned and what needs to be 'done' with that content);
– then create a learning environment that is likely to engage the student in activities that will bring about the intended outcomes;
– and finally use assessment tasks that directly address the intended outcomes and that enable assessors/teachers to judge if and how well students' performances meet the criteria.

As learning outcomes should be defined in relation with the curriculum of the students/pupils, the first step would be to explain how the outcome(s) fit(s) in with the curriculum. In my view, the lack of explicit verbalisation on how DDL activities meet curricular demands or expectations is one of the reasons for the lack of uptake of DDL in teaching contexts other than university-level courses (see for instance Boulton 2009; Gilquin, Granger 2010; Flowerdew 2012; Meunier 2020).

To provide a concrete example, the new 2017 curriculum for foreign language learning for the French-speaking-part of Belgium (Fédération Wallonie Bruxelles, FWB) explicitly mentions the importance of equipping learners with higher-level metacognitive processes. Learners are for instance expected to identify the linguistic resources that will be needed to carry out a specific task and to explain why these will be needed and how they can use them strategically.[2] In the structure of the observed learning outcomes (SOLO)

---

[2] E.g. excerpt from the document entitled "Compétences terminales et savoirs requis à l'issue des humanités générales et technologiques" (2017) produced by the Ministère de la Communauté Française, see http://www.enseignement.be/index.php?page=25189: «développer chez l'élève un niveau " méta " : être capable à la fois d'expliciter ses connaissances ou ses ressources, et de justifier les conditions dans lesquelles celles-ci peuvent être mobilisées. Il importe en effet de développer chez l'apprenant la conscience de ce que l'on peut faire de ses connaissances et compétences: 'je sais quand, pourquoi, comment utiliser tel savoir (concept, modèle, théorie…) ou tel savoir-faire (procédure,

taxonomy (Biggs, Collins 1989; **fig. 1**) the higher-level cognitive processes respectively include the ability:

– to analyse / apply / argue / compare / contrast / criticise / explain causes / relate / justify (relational cognitive processes)
– to create / formulate / generate / hypothesise / theorise, in our specific case, on and about language (extended abstract cognitive processes)

Most of those processes are at play when doing DDL activities (analysing, comparing, relating, generating hypotheses about language and the way it works).
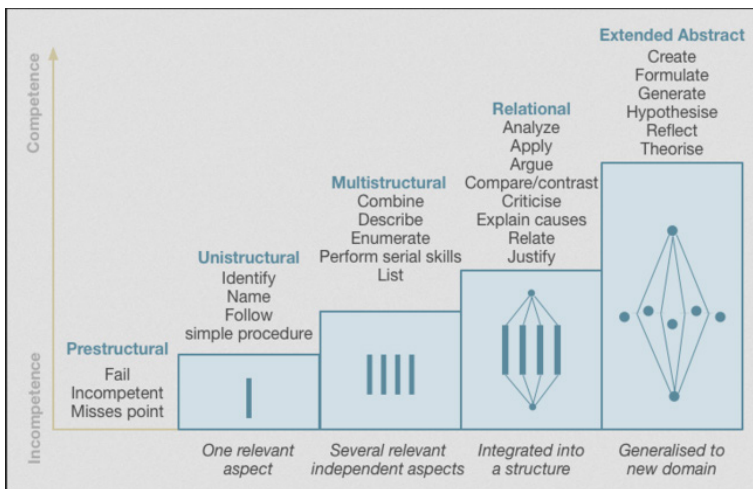

Figure 1 SOLO taxonmy (adapted from Biggs, Tang 2011 and available on the Open Source MOOC FutureLearn Course at https://www.futurelearn.com/courses/learning-teaching-university/0/steps/26410)

Once the need for acquiring higher-level cognitive skills has been established and clearly linked to the demands of the curriculum, it becomes easier to convince teachers and educational stakeholders of the value and benefits of including DDL activities in the teaching

démarche, stratégie...) ". Développer une telle capacité " méta " vise déjà un niveau de compétence relativement complexe» (p. 5); «démarche métacognitive (évaluation formative) de manière prospective et rétrospective : identifier les ressources linguistiques et les ressources stratégiques nécessaires et dire en quoi elles vont être nécessaires à la réalisation de la tâche en tenant compte des caractéristiques du type de production attendue, des éléments qui constitueront cette production» (p. 12).

activities.[3] It is also possible to select the type of DDL activity that best suits students/pupils in their specific contexts to meet their specific proficiency levels and needs. It is also important to scaffold the approach and gradually train students/pupils to become more autonomous.

Let's further exemplify alignment with a concrete linguistic focus. One of the learning outcomes can be that students should be able to use – certain types of – multiword units appropriately, as language is highly formulaic in nature (Wray 2002; Schmitt 2004). Using DDL activities seem particularly appropriate in such a case. One reason justifying the use of DDL is that the literature in Second Language Acquisition has shown that multiword units are subject to a number of interconnected determinants of learning (Ellis, Römer, O'Donnell 2016; Ellis 2017), including frequency effects, contingency (the association of forms and meanings) and various forms of learning (implicit vs. explicit). It has also been shown that multiword units are not easy to acquire for non-native speakers (contrary to what is the case for native speakers who acquire and use multiword units mostly implicitly). In addition, frequency information, prototypes and form-meaning mappings are much less directly accessible to non-native speakers and when alternative combinations exist only one is usually preferred (see Meunier 2012, 111). A second reason is that DDL relies on corpus technology which proves extremely useful as "EFL learners heavily depend on technology for learning authentic English" (Lui et al. 2014, 682), precisely because technology can help learners access, among other things, frequency information or form-meaning mappings.

Given what precedes we are in a situation where constructive alignment is taking shape:
– the foreign language curriculum recommends the need for training higher-level metacognitive processes,
– one of the identified difficulty in foreign language learning is the acquisition of multiword units,
– it is defined as one of the learning outcomes,
– we have access to one method (DDL) that not only makes it possible for students to access typical multiword units in the target language, but also does this by training them to become active constructors of knowledge by developing higher-level metacognitive skills,
– DDL should therefore be used regularly.

---

3 Some curricula do not mention the need for higher-level metacognitive skills. In such cases, it would be more difficult to argue for the inclusion of DDL activities.

Whilst this reasoning seems rather logical, there is still a notable lack of uptake of DDL activities. In the next section, I address some of the reasons for the current lack of sustainable practices in DDL in teaching contexts and suggest some paths towards possible improvements.

3 Reasons for the Current Lack of Sustainable Practices in DDL

Wilson (2013) states that corpora and DDL have failed to become established. My personal experience is also in line with that finding. I teach English for Academic Purposes classes at post-grad level and train my students in DDL to help them write their end-of-year paper. I am also involved in pre- and in-service teacher training sessions for upper secondary school teachers and, there too, I introduce DDL and try to present the advantages of the approach. Despite my motivation in trying to present all this in what I would define as a reasonably aligned way, I can only but notice that, in the two contexts, the initial interest does not transform into sustainable practices.

Wilson (2013, 32) points a series of reasons for this lack of uptake, among which:
- the technical complexity of corpus tools,
- the lack of promotion of corpora and DDL,
- the lack of collaboration between corpus developers and corpus users,
- a developer bias (tools not driven by needs of the end user but by those of the developer),
- a lack of supporting teaching materials and a suitable methodology for the application of DDL,
- the fact that many teaching materials reflect corpus linguists' interests but not users' needs.

In the rest of this section, I will focus on one of the issues listed above: the lack of promotion of corpora and DDL and argue for terminological simplification, especially when planning outreach activities for teachers. I will also comment on the lack of fully aligned practices, something that is not referred to by Wilson.

3.1 A Plea for Terminological Simplification

Whilst the use of terms such as corpus linguistics, DDL, formulaic language, metacognitive processes is probably acceptable in the present article, I have come to realise that the use of those terms can be detrimental in some contexts, particularly so in outreach activities

to teachers. Several discussions that I have had with teachers over the years proved extremely interesting and forced me to revise the terminology I tended to use spontaneously – and I am particularly thankful to the teachers who expressed their views in a constructive way. A term like 'corpus linguistics' is hardly known to teachers – despite the increasing number of textbooks being corpus-based. On average when I ask how many teachers have heard about corpora or corpus linguistics, about one tenth of the group has heard the term (which does not necessarily imply that they can explain it). The term 'linguistics' alone is perceived negatively by some teachers. Some of the teachers who had registered for a corpus linguistics seminar told me that they were "forced to register", "had no interested in linguistics" and were "looking for help with real problems their learners had". As for terms like formulaic language or metacognitive processes, they are often considered by teachers are academic jargon, which may create a feeling of distance

I would like to share some of minor terminological changes that I have recently adopted in in-service teacher training sessions and that have had major positive consequences. I no longer speak of DDL. Instead, I ask teachers (and sometimes learners too) if they would be happy to have a 24/7 native speaker assistant at their disposal. They are all enthusiastic about the offer and are a lot more active when it comes to using the tools and discovering what they can do with them.

Instead of using terms like formulaic language I show them how tools can highlight typical multiword units in a text. Instead of mentioning speech act formulae, I speak about everyday expressions such as 'You're welcome' or 'Take care'.

These little terminological changes have worked wonders in reducing teachers' anxiety and boosting their interest in using the tools. They also felt more at ease when introducing such tools to students as they did not feel the need to use complex terms.

I would thus recommend choosing the terms carefully when marketing corpora/DDL/formulaic language. As surprising as it may seem, not using the terms corpus linguistics or DDL actually improves their promotion to a wider audience. It is also essential to provide what Wilson (2013, 62) calls "aftercare". He argues that:

> It is one thing to organise expensive workshops with the aim of converting tutors to using corpora, but what happens after the workshop? What happens when a tutor encounters a problem or does not know how to perform a search? The likelihood is that he or she, without support, will lose interest in DDL and give it up. It is essential, therefore, that we identify ways of providing continuous support to corpus users. (62)

3.2   Constructive Alignment Includes Assessment

Another way of promoting sustainable DDL practices is to remember that assessment plays a central role in constructive alignment. In many cases, DDL activities are often considered as awareness raising activities and are only done on a few occasions. Little is done to check whether the awareness raising activity has had an impact and whether learners are able to use the acquired strategic and linguistic skills when necessary in their curriculum, for instance during a writing task. It is extremely rare to find students being allowed to consult corpora during a writing test. Reasons cited by teachers for not allowing corpus access to students include[4] technical problems (no easy computer access, no Wi-Fi), the fact that "they would no longer study their vocabulary", "it would be too easy", "they might cheat and access other websites".

Whilst some of these reasons cited are valid, I nonetheless believe that if one of the aims of our DDL teaching is to empower language learners with easier access to linguistic information otherwise difficult to learn/acquire through traditional types of learning (as is the case for collocations for example), then it would make sense to test whether students can make the most of the tools and methods we have spent time introducing to them. Allowing them to use corpora for specific writing tasks can go hand in hand with higher expectations in terms of accuracy or adherence to the features of a specific genre or text type. This will certainly not prevent students from studying their vocabulary in general, as they will probably also need to take oral tests where access to corpora will not be an option.

Including corpus literacy as one of the criterial features for some writing tasks is also one way of promoting sustainable corpus use. I asked my students (after analysing a questionnaire on their perception and actual use of DDL and corpus tools) why they often found corpus consultation and DDL useful but admitted not using such tools outside the classroom even for writing tasks. One of the most telling answers they gave me was that they "did not need" corpora except for my class, and that anyway they were never tested on their actual capacity to use the tools. I have since then changed the assessment setup and allow corpus consultation for a limited period of time during one of the specific writing task that is part of their final assessment. This has prompted motivation in my students to make sure they master the tools and are able to find the specific information they are looking for. They also now better understand the links between the learning outcomes of a course, the activities that they engage in to bring about the intended outcomes, and the assessment tasks that directly address the intended outcomes.

---

4   All the reasons listed have been provided by teachers during spoken interactions that took place in three teacher training seminar I gave at UCLouvain in 2016, 2017 and 2018.

4   Conclusion

In this article, I called for a more integrated and scaffolded approach to DDL. I presented a few concrete examples of how courses including DDL activities could be better aligned. I also discussed some of the reasons preventing sustainable DDL practices and proposed two solutions to increase sustainability, viz. a plea for terminological simplification in DDL, particularly when planning outreach activities for teachers and a revision of some language assessment practices.

To end this article I would like to highlight the importance of actively listening to teachers and learners that are being introduced to DDL and corpus tools. Their feedback has truly been instrumental in helping me adapt some of my teaching practices and, hopefully, in being slightly more successful in empowering them with new useful tools and practices for language learning.

The last word will be given to Wilson as I fully concur with him when he states that:

> the literature is proliferated by works extolling the benefits or advantages of DDL **over** other approaches. Surely it is more appropriate to consider the benefits or advantages of **combining** tradition with technology and introducing corpora to traditional teaching practices and/or to other innovative modes of delivery. (Wilson 2013, 65; emphasis added)

References

Biggs, J. (1996). "Enhancing Teaching through Constructive Alignment". *Higher Education*, 32(3), 347-64.

Biggs, J.; Collis, K. (1989). "Towards a Model of School-Based Curriculum Development and Assessment Using the SOLO Taxonomy". *Australian Journal of Education*, 33(2), 151-63. `https://doi.org/10.1177%2F168781408903300205`.

Biggs, J.; Tang, C. (2011). *Teaching for Quality Learning at University*. 4th ed. Berkshire: Open University Press.

Boulton, A. (2009). "Data-Driven Learning: On Paper, in Practice. Corpus Linguistics in Language Teaching". Harris, T.; Jaén, M.M. (eds), *Corpora in Language Teaching*. Bern: Peter Lang, 17-52. Linguistic Insights.

Boulton, A. (2017). "Data-Driven Learning and Language Pedagogy". Thorne, S.; May, S. (eds), *Language, Education and Technology*. New York: Springer. Encyclopedia of Language and Education, 181-92. `https://doi.org/10.1007/978–3–319–02237–6_15`.

Boulton, A.; Cobb, T. (2017). "Corpus Use in Language Learning: A Meta-Analysis". *Language Learning*, 67(2), 348-93. `https://doi.org/10.1111/lang.12224`.

Chambers, A. (2010). "What is Data-Driven Learning". O'Keeffe, A.; McCarthy, M. (eds), *The Routledge Handbook of Corpus Linguistics*. London: Routledge, 345-58. `https://doi.org/10.4324/9780203856949`.

Corino, E.; Onesti, C. (2019). "Data-Driven Learning: A Scaffolding Methodology for CLIL and LSP Teaching and Learning". *Frontiers in Education*, 4, art. 7. `http://doi.org/10.3389/feduc.2019.00007`.

Ellis, N.C. (2017). "Cognition, Corpora, and Computing: Triangulating Research in Usage-Based Language Learning". *Language Learning*, 67(1), 40-65. `https://doi.org/10.1111/lang.12215`.

Ellis, N.C.; Römer, U.; O'Donnell, M.B. (2016). *Usage-Based Approaches to Language Acquisition and Processing: Cognitive and Corpus Investigations of Construction Grammar*. New Jersey: Wiley-Blackwell. Language Learning Monograph Series.

Ellis, R. (1999). "Input-Based Approaches to Teaching Grammar: A Review of Classroom-Oriented Research". *Annual Review of Applied Linguistics*, 19, 64-80. `https://doi.org/10.1017/S0267190599190044`.

Flowerdew, L. (2012). *Corpora and Language Education*. Basingstoke: Palgrave Macmillan. Research and Practice in Applied Linguistics Series.

Forti, L. (2019). *Developing Phraseological Competence in Italian L2: A Study on the Effects of Data-Driven Learning* [unpublished PhD thesis]. Perugia: University for Foreigners of Perugia.

Gilquin, G.; Granger, S. (2010). "How can Data-Driven Learning be Used in Language Teaching?". O'Keeffe, A.; McCarthy, M. (eds), *The Routledge Handbook of Corpus Linguistics*. Routledge: London, 359-70.

Godwin-Jones, R. (2017). "Data-Informed Language Learning". *Language Learning Technology*, 21, 9-27. `https://doi.org/10.4324/9780203856949`.

Han, Z.; Park, E.S.; Combs, C. (2008). "Textual Enhancement of Input: Issues and Possibilities". *Applied Linguistics*, 29(4), 597-618. `https://doi.org/10.1093/applin/amn010`.

Johns, T. (1991). "Should You be Persuaded – Two Examples of Data Driven Learning Materials. Classroom Concordancing". *English Language Research Journal*, 4, 1-13.chambers.

Liu, S.; Liu, H.; Yu, Y.; Li, Y.; Wen, T. (2014). "TPACK: A New Dimension to EFL Teachers' PCK". *Journal of Education and Human Development*, 3(2), 681-93.

Meunier, F. (2012). "Formulaic Language and Language Teaching". *Annual Review of Applied Linguistics*, 32, 111-29. `https://doi.org/10.1017/S0267190512000128`.

Meunier, F. (2019). "A Case for Constructive Alignment in DDL: Rethinking Outcomes, Practices and Assessment in (Data-Driven) Language Learning". Crosthwaite, P. (ed.), *Data-Driven Learning for the Next Generation*. London: Routledge, 13-30. `https://doi.org/10.4324/9780429425899-2`.

Meunier, F. (2020). "Resources for Learning Multi-Word Items". Webb, S.A. (ed.), *The Routledge Handbook of Vocabulary Studies*. London: Routledge, 336-50. `https://doi.org/10.4324/9780429291586-22`.

Mishan, F.; Timmis, I. (2015). *Materials Development for TESOL*. Edinburgh: Edinburgh University Press.

Mizumoto, A.; Chujo, K. (2015). "A Meta-Analysis of Data-Driven Learning Approach in the Japanese EFL Classroom". *English Corpus Studies*, 22, 1-18.

Schmitt, N. (ed.) (2004). *Formulaic Sequences: Acquisition, Processing and Use*. Amsterdam: Benjamins.

Sharwood Smith, M. (1993). "Input Enhancement in Instructed SLA". *Studies in Second Language Acquisition*, 15(2), 165-79.

Vyatkina, N.; Boulton, A. (eds) (2017). "Corpora in Language Learning and Teaching", special issue, *Language Learning & Technology*, 21(3).

Wiggins, G.; McTighe, J. (2005). *Understanding by Design*. 2nd ed. Alexandria (VA): Association for Supervision and Curriculum Development ASCD.

Wilson, J. (2013). "Technology, Pedagogy and Promotion. How Can We Make the Most of Corpora and Data-Driven Learning (DDL) in Language Learning and Teaching?". *The Higher Education Academy*. `https://www.heacademy.ac.uk/system/files/corpus_technology_pedagogy_promotion2.pdf`.

Wray, A. (2002). *Formulaic Language and the Lexicon*. Cambridge: Cambridge University Press.