# Open science in second language acquisition research: The IRIS repository of research materials and data

*Emma* Marsden[1], *Sophie* Thompson[1], and *Luke* Plonsky[2]

[1]University of York, United Kingdom
[2]Georgetown University, United States of America

**Abstract.** IRIS (iris-database.org) is a freely accessible repository for materials, data collection tools, and data used in research on second language learning and teaching. Launched in 2011, IRIS now holds nearly 4,000 files covering a vast spectrum of theoretical, methodological, and epistemological perspectives. The current chapter outlines the background, purpose, content, and usage of IRIS both generally and as within the context of the current volume. We also describe the place of IRIS as one of the primary vehicles driving the move toward open science practices and replication research in applied linguistics. We conclude by encouraging engagement with IRIS from research and practitioners alike.

**Résumé.** IRIS (iris-database.org) est un entrepôt en accès libre pour du matériel, des instruments de collecte de données et des données utilisées dans des recherches sur l'apprentissage et l'enseignement d'une langue seconde. Inauguré en 2011, IRIS contient à présent près de 4000 fichiers qui couvrent un vaste spectre de perspectives théoriques, méthodologiques et épistémologiques. Ce chapitre présente le contexte, l'objectif, le contenu et l'usage d'IRIS d'une façon générale ainsi que dans le cadre du présent volume. Nous y décrivons le rôle que joue IRIS comme un des vecteurs principaux du mouvement vers des pratiques scientifiques ouvertes et une recherche réplicative en linguistique appliquée. Nous concluons en encourageant les chercheurs et les praticiens à collaborer avec IRIS.

## 1 Introduction

We welcome this volume of research unified by cognitive and usage-based approaches to understanding second language learning and teaching. The papers also provide a cohesive and much-needed focus on, mainly, the acquisition of French as a second language (L2). The studies provide many useful insights into a wide range of sub-domains of L2 research, including the effects of explicit knowledge on brain activity, the role of interaction, relations between conceptual and lexico-semantic representations, and the development of L2 pragmatics, phonology and morphosyntax. The methods used to elicit data for these diverse

topics are, inevitably, similarly diverse. We also applaud the editors and authors for making their work open access, thus embracing current moves towards open science practices.

In these ways, the volume aligns itself very well with the aims and content of the IRIS (Instruments for Research into Second Languages) repository (Marsden, Mackey, & Plonsky, 2016). Like the volume, IRIS also embraces a wide diversity of substantive areas and methodological approaches (see for example Mackey & Marsden 2016), and it is at the forefront of open science practices. In the current chapter, we describe the background, purpose, content, and usage of IRIS. We also discuss one of the long-term benefits of our field embracing open science practices: improvements in the quality and quantity of replication research.

Thus, our chapter extends the brief discussion of IRIS that followed Marsden's keynote talk at the COULS (2016) conference. The studies presented in that talk have their materials openly available on IRIS: Kasprowicz & Marsden (2017) (L2 German morphosyntax markers of subject/object roles); Marsden (2006) (L2 French inflectional morphology for person, number and tense); Marsden & Chen (2011) (L2 English inflectional morphology for tense); Marsden, Williams, & Liu (2013) (suffixes for tense and number in a novel language); McManus & Marsden (2017 & in press) (French inflectional morphology for aspect in the imparfait).

## 2 Background

IRIS is a digital repository of materials used to collect data for research into second and foreign language learning and teaching. Since its launch in 2011, with initial support from the UK's *Economic and Social Research Council* followed by long-term support from the *British Academy*, IRIS has attracted increasing international interest, with around 38,000 hits and 20,000 downloads. The main purpose of the repository is to make instruments and data from second language learning and teaching research and multilingualism openly accessible for researchers, teachers and students. IRIS has the support of a wide network of leading journal editors and research and teaching associations.

## 3 Content of the IRIS repository

IRIS contains more than 3,600 materials, with an increase of 1,500 materials in the last 12 months. These include different kinds of tests for a wide range language features in a variety of languages, such as, judgement tests, elicited imitation tasks, stimuli for psycholinguistic techniques, teaching intervention activities, questionnaires, interview protocols, observation schemes, software scripts, sound and video files, among many more types of materials. IRIS is searchable by many different parameters, including: the kind of material (e.g., proficiency test, questionnaire); the area of research being conducted (e.g. pragmatics, inter-cultural competence, morphosyntax, individual differences); the names of researchers who contributed or developed the materials; learner characteristics (e.g. age, first language, proficiency); the language being learned; and the type of instruction or pedagogical technique (e.g. processing instruction, corrective feedback). The materials available on IRIS are free to download, can be adapted to suit the downloaders' own requirements, and have all been used in peer-reviewed publications or successful PhD theses in the broad area of second language learning and teaching and multilingualism. IRIS provides access to full materials, which is critical given that space in journal articles and books is often limited and reporting practices are known to omit details about instrumentation (Derrick, 2016).

In addition to increasing the number of individual instruments on IRIS over the last year, IRIS has developed special collections holding as many examples as possible of a specific

type of research instrument. Currently two collections are held: <u>self-paced reading</u> (SPR) tests and <u>grammaticality/acceptability judgement tests (JT)</u>. Prior to these special collections on IRIS, the availability of these instruments was lacking. Full instruments were available - and mainly behind journal paywalls - for only 33% of SPR studies and 34% JT studies in the field (as found by Marsden, Thompson and Plonsky, under review; Plonsky, Marsden, Crowther, Spinner, Gass, in preparation). There are now many more complete SPRs and JTs available, accessible via IRIS. The development of more special collections is an important goal for IRIS.

# 4 Usage

IRIS is used by researchers (16% of downloads), teachers (17% of downloads), and students (67% of downloads). Researchers can use IRIS to share and archive their materials, to increase the impact and exposure of their work, and to scrutinise, replicate and build on the work of their colleagues. Graduate students can adapt instruments to investigate a wide range of areas in TESOL and applied linguistics, for example, cross-linguistic influence, sentence processing, testing, to name but a few. Research methods trainers use IRIS to help students to understand and critique methodologies and to help them when developing their own materials for dissertations. Language teachers are able to use or adapt materials from IRIS, such as tests for assessing learners' needs, intervention materials for teaching activities, or instruments they can use for their own research to understand more about how to give corrective feedback or about how learners are motivated.

# 5 Wider purpose and context: Open science and replication

Beyond simply making materials available to second language teachers and researchers, IRIS is at the forefront of the open science movement currently taking shape in applied linguistics and across the social and natural sciences. 36 top journals in the field support IRIS by encouraging their authors to share materials through the repository. Leading journals (*Language Learning, Studies in Second Language Acquisition, The Modern Language Journal*, with others set to follow) now do this by awarding "open science badges" to researchers who openly share their materials or data, in collaboration with the *Centre for Open Science* (see Figure 1). This level of activity shows the increasing use and impact of IRIS in language learning and teaching research. It is also a sign of the field's growing recognition of the place and value of open science practices.



**Fig.1.** Badges awarded by the Center for Open Science to recognise open science practices

Another wider benefit of IRIS is that it can assist the field in overcoming the lack of replication studies on language learning and teaching. Replication research allows us to investigate the external validity (generalizability) of findings in, for example, different settings, demographics, target languages, helping researchers to test theories, constructs, and

measurements (Marsden, Morgan-Short, Thompson, & Abugaber, 2018; Porte, 2012). It can also inform us about the reliability of research and of measures with the same or different samples of learners (Plonsky & Derrick, 2016). IRIS is actively facilitating and encouraging replication through the growing database of materials and through the new IRIS replication award. The award recognizes researchers who publish rigorous replication research using materials from IRIS, and, critically, self-label their study as a 'replication'. Self-labelling a study as a 'replication' allows reviewers and readers to track methodological and theoretical precedents of research, helping us to understand more precisely the extent to which individual studies build on and inform previous studies. A very small amount of self-labelled replication in L2 research has been published to date, with an estimated rate of less than 0.26, that is, fewer than one in every four hundred published articles is a replication study (Marsden et al., 2018). The IRIS award is intended to encourage researchers to carry out closer replications, and to feel confident in self-labelling them as such (e.g. Morgan-Short, Marsden, Heil, et al. 2018).

## 6 Concluding remarks

In summary, IRIS is a user-friendly, searchable, easily-accessible database which can be used by all those involved in second language research. In collaboration with COS, journals supporting open science, and the research community IRIS is furthering the open science agenda and facilitating research rigour.

This volume has led to several useful contributions to the IRIS database. For example, researchers can now easily access, and adapt to their own contexts:

a new discourse completion test on requests in Finnish (Holttinen),

an oral production task on counterfactuality (Repiso)

a phonetic identification task (Maurova Paillereau)

transcription data (Chachu)

We hope that the availability of these instruments will lead to their further adaptation and use in the future.

For more information, please see the FAQ on the IRIS website or contact iris@iris-database.org. Follow us on Facebook: https://www.facebook.com/irisdatabase/

## References

Derrick, D. J. (2016). Instrument reporting practices in second language research. *TESOL Quarterly, 50,* 132-153.

Kasprowicz, R. E., & Marsden, E. (2017). Towards ecological validity in research into input-based practice: Form spotting can be as beneficial as form-meaning practice. *Applied Linguistics*, 1-27. DOI: 10.1093/applin/amw051

Mackey, A. & Marsden, E. (Eds.) (2016). *Advancing Methodology and Practice: The IRIS Repository of Instruments for Research into Second Languages*. New York: Routledge.

Marsden, E. (2006). Exploring Input Processing in the classroom: An experimental comparison of Processing Instruction and Enriched Input, *Language Learning*, *56,* 507-566

Marsden, E. & Chen, H.-Y. (2011). The roles of structured input activities in processing instruction and the kinds of knowledge they promote. *Language Learning, 61,* 4, 1058–1098.

Marsden, E., Mackey A., Plonsky, L. (2016). The IRIS repository: Advancing research practice and methodology. In Mackey, A. & Marsden E. (Eds.) *Instruments for Research into Second Languages: Empirical studies advancing methodology* (pp. 1-21). New York: Routledge.

Marsden, E. & Plonsky, L. (2018). Methodological transparency in data collection and instrument design. In A. Edmonds & A. Gudmestad (Eds) *Critical reflections on data in SLA.* Amsterdam: John Benjamins.

Marsden, E., Thompson, S., Plonsky, L. (under review). Self-paced reading in second language research: A methodological synthesis and recommendations for the field. *Applied Psycholinguistics*

Marsden, E., Williams, J., Liu, X. (2013). Learning novel morphology: The role of meaning and orientation of attention at initial exposure. *Studies in Second Language Acquisition*, *35*, 4, 619-654.

McManus, K. & Marsden E. (2017). Explicit instruction about the first language can benefit online and offline performance in a second language. *Studies in Second Language Acquisition*, *39*, 3, 459-492.

McManus, K., & Marsden, E. J. (in press). Online and offline effects of L1 practice in L2 grammar learning: a partial replication. *Studies in Second Language Acquisition*.

Marsden, E., Morgan-Short, K., Thompson, S., Abugaber, D. (2018) Replication in second language learning research: A narrative and a systematic synthesis and lessons for the field. *Language Learning 68*, 1

Morgan-Short, K., Marsden, E., Heil, J. et al. (under review). Attention to form and comprehension: A multi-site replication. *Language Learning,* *68*, 1

Plonsky, L., & Derrick, D. J. (2016). A Meta-Analysis of Reliability Coefficients in Second Language Research. *The Modern Language Journal*, *100,* 2, 538-553.

Porte, G. (2012). *Replication research in applied linguistics*. New York: Cambridge University Press.